Staphylococcus aureus virulence gene studies : a comparative microarray based approach

Thesis

STAPHYLOCOCCUS AUREUS VIRULENCE GENE STUDIES: A COMPARATIVE MICROARRAY BASED APPROACH

A thesis submitted in fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

DEQA HASSAN A. MOHAMED (BSc, MRes)

November 2009

This research project was carried out in collaboration with the

Health Protection Agency

Centre for Infections

61 Colindale Avenue, London NW9 5HT

United Kingdom

THESIS CONTAINS

CD

ABSTRACT

The development and application of a partial composite *S. aureus* virulence-associated gene microarray is described. Epidemic, pandemic and sporadic lineages of healthcare-associated (HA-) and community-associated (CA-) *S. aureus* were compared. The clonal population structure was supported but further evidence for large-scale recombination events was obtained. Phage structural genes linked with the CA phenotype were identified and *in silico* analysis revealed these to be correlated with phage serogroup. CA strains generally carried a PVL-associated phage either of the A or Fb serogroup, whilst the HA strains predominantly carried serogroup B phage. It is proposed that carriage of PVL-associated phage rather than the specific *pvl* genes is correlated with the CA phenotype. These findings further support the role of the accessory genome in shaping the epidemiology of *S. aureus*.

The microarray was used to study gene expression in isogenic strains differing by a deletion in the *agr* locus. Microarray analysis revealed significant differences between the levels of expression of several genes of the normal and mutant strains. However, RNAIII levels in the non-mutant strain were found to be cell density independent, indicating that the expected quorum sensing mechanism was not functional.

Expression profiles of cells grown under biofilm simulating conditions were compared to their planktonic counterparts. Biofilm cells displayed a typical expression profile that was different from both the actively growing planktonic exponential cells and the planktonic stationary cells. The strongest feature of the biofilm state was high level expression of the haemolysin genes. This model therefore is amenable to exploitation in studies designed to improve our understanding of the mechanisms underlining biofilm survival and regulation after long periods of growth.

AUTHORS DECLARATION

No part of this thesis has been submitted in support of an application for any degree or qualification of any other university or institute of learning. The work of this thesis was carried out by myself (unless otherwise stated in the text), under the supervision of Professor Nick Saunders of the Health Protection Agency.

Deqa Hassan A. Mohamed

November 2009

ACKNOWLEDGEMENTS

I am most grateful to my principal supervisor Professor Nick Saunders for his continuous supervision, advice, guidance, encouragement and support. Without his help and constructive criticism, this study would not have been possible.

I extend my gratitude to Dr Angela Kearns (my second supervisor) for her guidance, encouragement, advice and helpful discussions. Her expertise has been invaluable in shaping the writing of this thesis. Thanks also to Dr Kearns and my colleagues in LHCAI (Colindale) for the provision of the *Staphylococcus aureus* strains and the molecular typing data used in this study.

I thank my sponsors, NACBO (Novel and Improved Nanomaterials, Chemistries and Apparatus for Nano-Biotechnology; grant number 500804-2), for providing financial support for this study, and supporting my attendance to national and international conferences.

I wish to thank my colleagues Dr Sayeh Sabersheikh, Dr Craig Swift and Dr Peter Staves for moral support, advice and encouragement. Many thanks are also due to my colleagues and friends at the HPA who made my stay at Colindale so enjoyable.

Special thanks to my sisters, husband and my wonderful family for their patience, endless encouragement and unwavering support. Without their encouragement, this work could not have been accomplished.

Finally, to my parents, who taught me that a good education lasts a lifetime. My success, is your success. This work is dedicated, with love, to you.

TABLE OF CONTENTS

(Page Numbers)

LIST OF TABLES

(Page numbers)

LIST OF FIGURES

CHAPTER 1.0 INTRODUCTION

1.1 THE STAPHYLOCOCCI

Staphylococci belong to the Order Bacillales, Bacterial family Staphylococcaceae, genus *Staphylococcus*. They are characteristically non-motile, Gram positive, non sporulating facultative anaerobes, which generate energy from respiration and fermentation. They are also catalase positive. Variations in colony pigmentation are common, ranging from opaque to orange (Holt, 1994). The genus *Staphylococcus* has been studied extensively and contains at least 41 recognised species and 24 subspecies (January 2009 entry at http://www.bacterio.cict.fr/s/staphylococcus.html). The genus can also be sub divided into two broad groups based on ability to coagulate blood plasma. Coagulase-positive species are primarily represented by *S. aureus*, but also include *S. intermedius* and *S. delphini*. Coagulase-negative species include *S. epidermidis* (Kloos, 1990).

Based on their ability to divide in several planes, *S. aureus* display a cluster appearance on microscopic examination; a property first-described by Sir Alexander Ogston in 1880. Ogston observed these organisms in pus from human abscesses. On microscopic examination, the appearance of the organisms was described as similar to grapes (Greek name *staphyle*) and berries (*kokkos*). Shortly after in 1884, staphylococci isolated from a wound were grown for the first time in pure culture by Rosenbach (Baird-Parker, 1990).

Staphylococci form part of the normal flora of the human nasopharynx, other musocal surfaces and skin. Infections associated with staphylococci are generally the result of breaches in the host's innate immunity, such as damage to mucosal and cutaneous membranes (through trauma, surgery or indwelling medical devices) (Lowy, 1998). It is for this reason staphylococci are referred to as opportunistic pathogens (Kloos 1980,

Massey *et al.* 2006). *S. aureus* is the most clinically significant member of this genus (Holt, 1994) and has been studied extensively. It can be distinguished from other species based on its gold colony pigmentation, positive coagulase reaction, ability to ferment mannitol and deoxyribonuclease activity. *S. epidermidis* (the most common species of coagulase-negative staphylococci from clinical disease), is responsible for infections associated with medical devices e.g. implanted prosthetics or intravascular devices (von Eiff, 2002). Prior to the use of these devices in modern medicine, *S. epidermidis* was rarely considered pathogenic. Conversely, *S. aureus* is more aggressive in causing infection, both acute and chronic in nature.

1.2 THE STAPHYLOCOCCUS GENUS

1.2.1 *Staphylococcus aureus*

S. aureus is a highly adapted and extremely successful coloniser of the moist squamous epithelium of the anterior nares in humans (30-70% of the population), other mucosal surfaces and skin. It is also occasionally found as part of the flora of the digestive and vaginal tracts (Smith *et al.* 1982, Enright 2008). The innate and adaptive components of the nasal associated lymphoid tissue (NALT) protect the nares from microbial infection. NALT organs (tonsils and adenoids) and mucosal membranes both consist of single-layers of epithelial cells which are continually washed in mucus. Mucus contains further protective elements including antimicrobial proteins and peptides, proteoglycans, specialized immune cells and secretory immunoglobulins (Fokkens and Scheeren 2000, Massey *et al.* 2006). *S. aureus* is able to maintain its colonist status by producing a variety of virulence-associated factors that interact with host cell components. Several of the virulence factors described for *S. aureus* permit immune evasion whilst others are toxins enabling the spread of infection (discussed further in section 1.3). These virulence

factors facilitate its survival as a commensal in this niche. However, *S. aureus* is also able to infect any tissue of the body via a breach in the skin or mucous membranes. Infections which then develop locally can potentially disseminate; diseases range in severity from minor skin infections to systemic life threatening conditions. Most individuals (if not all) will succumb to infection with *S. aureus* during their lifetime. The most common of the uncomplicated infections include pimples, boils, styes and conjunctivitis. More severe infections include endocarditis, osteomyelitis, haemorrhagic pneumonia and other metastatic complications. *S. aureus* has also been noted for causing toxin-mediated diseases such as toxic-shock syndrome, food poisoning and scalded skin syndrome (Lindsay *et al.* 2004, Kuroda *et al.* 2001). However, invasive disease and infection of normally sterile body sites are very rare in previously healthy individuals (Enright, 2008).

1.2.1.1 *S. aureus* genomics

S. aureus is unsurpassed among human pathogens in the versatility of its pathogenic strategies, number of virulence factors, and ability to survive and multiply in a wide range of environments (Oliveira *et al.*, 2002). The impact on human health of *S. aureus* infections in community and hospital settings has led to intensive investigation of this organism over recent years (Holden *et al.*, 2004). Incremental changes in the gene complement have resulted in the emergence of strains that are antibiotic-resistant, transmissible and successful in causing disease (Holden *et al.*, 2004).

With the advent of molecular biology techniques, a greater understanding of the mechanisms governing these aspects has been elucidated. In particular, genome sequencing in recent years has advanced our understanding of the biology and genetics of this species. To date, 19 staphylococcal genomes (NCBI) have been sequenced (publicly available), of which 14 are *S. aureus* genomes (Table 1.1). The main aim of these projects

has been to understand the nature of staphylococcal pathogenicity, and why certain strains/species are more successful than others. In general, these initiatives have shown the genomes of staphylococci share a common core set of genes, in addition to accessory genes. In *S. aureus*, these accessory components are horizontally acquired and are composed of genomic islands, transposons, plasmids, insertion sequences and bacteriophage-derived segments. Encoded on these components are many virulence factors and antibiotic resistance genes thought to contribute to the pathogenic potential of this organism. In particular, a prominent concern with respect to *S. aureus* is the rate at which it acquires antibiotic resistance, due to the accumulation of mobile genetic elements encoding resistance genes from other species (or within the species). The complexity of the regulation of virulence factors (discussed further in section 1.3) provides *S. aureus* with the adaptability to interact with its host in such a diverse manner to colonise, spread and cause disease (whether toxin-mediated, pyogenic or invasive).

Table 1.1 The fourteen completely sequenced *Staphylococcus aureus* strains.

Strain	CC*	Source	First isolated	Comments*
N315	5	Pharynx, Japan	1982	Hospital-acquired MRSA.
Mu50	5	Wound, Japan	1997	Hospital-acquired VISA, related to N315.
MW2	1	Fatal paediatric bacteraemia, North Dakota, USA	1998	Typical USA community-acquired MRSA, PVL-positive.
MRSA252	30	Fatal bacteraemia, Oxford, UK	1997	Typical UK hospital-acquired epidemic MRSA (EMRSA-16).
MSSA476	1	Osteomyelitis, Oxford, UK	1998	Community-acquired MSSA.
COL	8	Colindale, UK	1961	Early MRSA.
NCTC8325	8	Colindale, UK	<1949	Laboratory strain, parent of non-lysogenic 8325-4, can be genetically manipulated.
USA300-FPR3757	8	USA	2000	Community associated MRSA, PVL-positive.
USA300-TCH1516	8	Severe sepsis, USA	2007	Community associated MRSA, PVL-positive.
RF122	151	Ireland	2006	Associated with bovine mastitis
Mu3	5	Pneumonia, Japan	1996	Hetero-VISA
JH1	5	Endocarditis, USA	2000	MRSA
JH9	5	Endocarditis, USA	2000	*In vivo* VISA derivative of JH1
Newman	8	UK	<1941	MSSA, clumping factor over-producer

*Abbreviations: CC, clonal complex; MRSA, methicillin-resistant *S. aureus*; MSSA, methicillin-sensitive *S. aureus;* PVL, Panton-Valentine leukocidin; VISA, vancomycin intermediate-level-resistant *S. aureus*. (Adapted from Lindsay and Holden., 2004, Holden and Lindsay 2008)

1.2.1.2 Antimicrobial resistance in *S. aureus*

1.2.1.2.1 Molecular basis of antimicrobial resistance

S. aureus is a leading cause of nosocomial infections and it is also becoming of increasing concern in the community. This is due to its transmissibility, intrinsic virulence, ability to cause a diverse array of life-threatening infections, and capacity to adapt to different environmental conditions (Lowy, 2003). The vast genetic repertoire of this bacterium allows it to adapt rapidly to change in hostile environments. This ability is repeatedly demonstrated by the emergence of strains that have acquired resistance to virtually all antimicrobial agents shortly after their introduction into clinical practice.

The introduction of large numbers of structurally diverse antimicrobial agents into human use over the past 60 years has presented a new set of challenges to bacterial pathogens such as *S. aureus*. Successful lineages of contemporary pathogens have excelled in their ability to acquire resistance genes and to construct regulatory mechanisms that can survive following exposure to increasing concentrations of the antimicrobial agent. Furthermore, these resistant strains have the propensity to spread, establish ecological reservoirs, colonise, and cause disease (Oliveira *et al.*, 2002). The ever increasing emergence of antimicrobial resistant bacterial species can be attributed to a multitude of factors that include the widespread and sometimes inappropriate use of antimicrobials, their extensive use as growth enhancers in animal feed, an increase in regional and international travel and the relative ease with which these organisms cross geographic barriers (Lowy, 2003).

In the post-antibiotic era, staphylococci have developed efficient mechanisms to neutralise each new antibiotic introduced. Antimicrobial resistance can occur by the mutation of

chromosomal loci (e.g. for streptomycin, rifampin, fusidic acid, mupirocin and novobiocin) (Wichelhaus *et al.* 2001, O'Neil *et al.* 2006, Vickers *et al.* 2007). However, the primary mechanism driving multiple antimicrobial resistance in staphylococci is through horizontal gene transfer of pre-formed resistance determinants carried on accessory genetic elements including plasmids, transposable elements (insertion sequences and transposons) and genomic islands (Lyon and Skurray 1987, Jensen and Lyon 2009). Understanding the mechanisms by which *S. aureus* resists antimicrobials is a pre-requisite to developing more robust therapeutics. Such research has highlighted a few key mechanisms used by *S. aureus* (Table 1.2; reproduced from Lowy 2003). A recent review (Jensen and Lyon, 2009) highlights several examples of antimicrobial resistance determinants in *S. aureus* and the mechanisms of resistance.

Table 1.2 Mechanisms of *S. aureus* resistance to antimicrobials. Taken from Lowy, 2003.

Antibiotic	Resistance gene(s)	Gene product(s)	Mechanism(s) of resistance	Location(s)
β-Lactams	1) *blaZ*	1) β-Lactamase	1) Enzymatic hydrolysis of β-lactam nucleus	1) Pl:Tn
	2) *mecA*	2) PBP2a	2) Reduced affinity for PBP	2) C:SCC*mec*
Glycopeptides	1) Unknown (VISA)	1) Altered peptidoglycan	1) Trapping of vancomycin in the cell wall	1) C
	2)	2) D-Ala-D-Lac	2) Synthesis of dipeptide with reduced affinity for vancomycin	2) Pl:Tn
Quinolones	1) *parC*	1) ParC (or GrlA) component of topoisomerase IV	1,2) Mutations in the QRDR region, reducing affinity of enzyme-DNA complex for quinolones	1) C
	2) *gyrA* or *gyrB*	2) GyrA or GyrB components of gyrase		2) C
Aminoglycosides (e.g., gentamicin)	Aminoglycoside-modifying enzymes (e.g., *aac*, *aph*)	Acetyltransferase, phosphotransferase	Acetylating and/or phosphorylating enzymes modify aminoglycosides	Pl, Pl:Tn
Trimethoprim-sulfamethoxazole (TMP-SMZ)	1) Sulfonamide: *sulA*	1) Dihydropteroate synthase	1) Overproduction of *p*-aminobenzoic acid by enzyme	1) C
	2) TMP: *dfrB*	2) Dihydrofolate reductase (DHFR)	2) Reduced affinity for DHFR	2) C
Oxazolidinones	*rrn*	23S rRNA	Mutations in domain V of 23S rRNA component of the 50S ribosome. Interferes with ribosomal binding	C
Quinupristin-dalfopristin (Q-D)	1) Q: *ermA*, *ermB*, *ermC*	1) Ribosomal methylases	1) Reduce binding to the 23S ribosomal subunit	1) Pl, C
	2) D: *vat*, *vatB*	2) Acetyltransferases	2) Enzymatic modification of dalfopristin	2) Pl

[A]Examples of several of the *S. aureus* mechanisms of resistance to selected antibiotics
Pl, plasmid; C, chromosome; Tn, transposon; QRDR, quinolone resistance-determining region.

1.2.1.2.2 Resistance to β-lactam antimicrobials

1.2.1.2.2.1 Penicillin resistance

Penicillin, the first β-lactam antibimicrobial, was introduced in the early 1940s. Prior to its introduction, the mortality of patients with *S. aureus* bacteraemia was greater than 80% (Skinner and Keefer, 1941) and over 70% developed metastatic infections. The introduction of penicillin heralded a dramatic improvement in prognosis for these patients. However, by the late 1960s, greater than 80% of both hospital- and community-associated staphylococci were penicillin-resistant (Lowy, 2003).

The mechanism of staphylococcal resistance to penicillin was later shown to be due to the effects of the β-lactamase enzyme. Staphylococci synthesize β-lactamase when exposed to β-lactam antibiotics (fig. 1.1; reproduced from Lowy 2003). Synthesis of β-lactamase permits inactivation of penicillin (and other β-lactam antimicrobials e.g. ampicillin and amoxicillin) via hydrolysis of the β-lactam ring. The gene that encodes β-lactamase, *blaZ*, is part of a transposable element located on a large plasmid, which often carries additional antimicrobial resistance genes (e.g. to erythromycin and gentamicin) (Lowy, 2003).

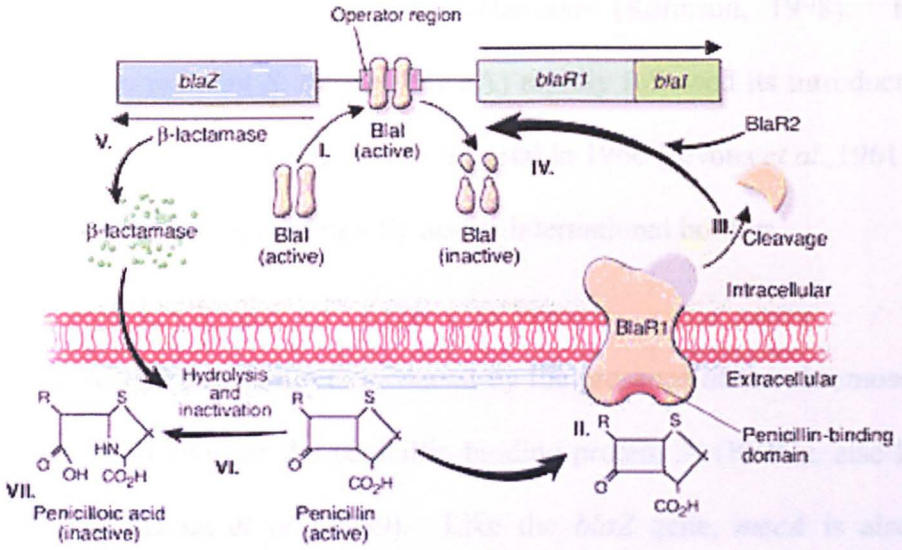


Figure 1.1 Induction of staphylococcal β-lactamase synthesis in the presence of the β-lactam antibiotic penicillin. The *blaZ* gene is controlled by two adjacent regulatory genes, the anti-repressor *blaR1* and the repressor *blaI.* (I) The DNA-binding protein BlaI binds to the operator region, thus repressing RNA transcription from both *blaZ* and *blaR1-blaI.* In the absence of penicillin, β-lactamase is expressed at low levels. (II) Binding of penicillin to the transmembrane sensor-transducer BlaR1 stimulates BlaR1 autocatalytic activation. (III–IV) Active BlaR1 either directly or indirectly (via a second protein, BlaR2) cleaves BlaI into inactive fragments, allowing transcription of both *blaZ* and *blaR1-blaI* to commence. V–VII. β-Lactamase, the extracellular enzyme encoded by *blaZ* (V), hydrolyzes the β-lactam ring of penicillin (VI), thereby rendering it inactive (VII). Taken from Lowy, 2003.

1.2.1.2.2.2 Methicillin resistance

Methicillin (or meticillin), a penicillin derivative introduced in 1959, was the first of the semi-synthetic penicillinase-resistant penicillins. Originally called celbenine, methicillin was the first mechanism-based antimicrobial agent. As a semi-synthetic derivative of penicillin, it was chemically modified to withstand the degradative action of penicillinase (Oliveira *et al.*, 2002). The introduction of methicillin and other semi-synthetic penicillins (e.g. oxacillin, cloxacillin, flucloxacillin) and penicillinase-resistant methicillin represented a significant advancement in anti-staphylococcal therapy; these agents were non

susceptible to the activity of β-lactamases (Rolinson, 1998). However, reports of methicillin-resistant *S. aureus* (MRSA) rapidly followed its introduction. The first report of resistance occurred in a British hospital in 1960 (Jevons *et al.*,1961). Successful MRSA clones have since spread rapidly across international borders.

Resistance to methicillin is mediated by the presence of the chromosomally located *mecA* gene. *mecA* encodes the penicillin-binding protein 2a (PBP2a, also known as PBP2' and *mecA*) (Ubukata *et al.*, 1989). Like the *blaZ* gene, *mecA* is also controlled by two regulatory proteins, MecR1 and MecI. In the presence of β-lactam antibiotics, MecR1 protein (signal transducer) directly or indirectly cleaves the MecI repressor which is bound to the *mecA* operator. Cleavage of MecI allows transcription of *mecA* to occur, and thus synthesis of PBP2a (Katayama *et al.*, 2001). PBP2a has been shown to have a reduced affinity for methicillin, and was proposed to permit cell wall synthesis (normally blocked by β-lactams) in the presence of β-lactam antibiotics. Synthesis of the peptidoglycan layer is a critical function necessary for cell homeostasis and thus growth of MRSA (Hartman and Tomasz 1984, Chambers *et al.* 1985, Lowy 1998, Berger-Bachi and Roher 2002).

Although *mecA* confers resistance to most β-lactam antibiotics, not all *mecA*-positive strains are equally resistant to methicillin. The overall level of resistance in an MRSA population depends on the interplay between several chromosomal determinants affecting the expression of PBP2a. For this reason, MRSA resistance levels vary from phenotypically susceptible (MICs as low as 1μg/ml) to highly resistant (MIC >500μg/ml) (Berger-Bachi and Rohrer, 2002). The genomic factors controlling resistance levels, termed fem/faux factors, include genes involved in peptidoglycan precursor formation, composition and turnover, teichoic acid synthesis, and several genes of uncharacterised (or poorly characterised) function (de Lencastre *et al.* 1994, de Lencastre *et al.* 1999; Berger-Bachi and Rohrer 2002, Rohrer and Berger-Bachi 2003). Several regulatory loci have also

been shown to influence resistance levels. These include regulators of metabolism e.g. catabolite control protein A (CcpA) (Seidl *et al.,* 2006), global regulators of virulence factors such as the *agr* and *sarA* systems and the alternative sigma factor σ^B (Piriz Duran *et al.* 1996, Wu *et al.* 1996) and the VraSR loci, responsible for cell wall stress response induced upon antibiotic challenge (Kuroda *et al.,* 2003). SA1665, a recently characterised DNA binding protein, is amongst the latest implicated in modulating β-lactam resistance (Ender *et al.,* 2009). SA1665, which has a negative impact on methicillin resistance, has been proposed to modulate β-lactam resistance in a *mecA*-independent manner by controlling genomic factors or cellular functions necessary for methicillin resistance (Ender *et al.,* 2009).

Studies on the evolution of methicillin resistance in *S. aureus* have shown that this occurred in a stepwise manner. Pre-MRSA strains carried the *mecA* gene along with the regulatory gene *mecI* and *mecR1*, but displayed no resistance. Hetero-MRSA strains displayed a mutation in *mecI* causing release of repression effects of *mecI*; strains then showed low level resistance to minimal concentrations of methicillin but remained susceptible to high dosages. Subsequently, homo-MRSA displayed homogeneously high resistance to methicillin (Hiramatsu, 2004).

MRSA have been noted for their accumulation of resistance determinants within the *mec* region (section 1.2.1.3.2), an integration hotspot for plasmids and transposons, thus giving rise to multi-drug resistance (Ito *et al.,* 1999). It is this ability of *S. aureus* to rapidly acquire resistance to antibiotics that has led to its classification as a major pathogen or "superbug".

1.2.1.2.2.3 Vancomycin resistance

MRSA infections are generally treated with glycopeptides including vancomycin. Vancomycin is frequently administered for empirical therapy in patients displaying signs of infection, prior to MRSA diagnosis that can take up to 48 hours. Increased use of vancomycin has led to the emergence of two types of glycopeptides-resistant *S. aureus*; vancomycin intermediate-resistant *S. aureus* (VISA) and vancomycin-resistant *S. aureus* (VRSA). The first VISA (Mu50), isolated in Japan, was described in 1997, and has since spread world-wide (Tenover, 1998). Glycopeptides inhibit cell wall synthesis in Gram positive bacteria. VISA strains synthesise a thicker cell wall that absorbs the vancomycin and allows sufficient cross-linking of newly synthesised cell wall (Pereira *et al.*, 2007). VISA strains are generally unfit and slow-growing, arising through spontaneous mutations in patients undergoing long-term glycopeptide therapy (Mwangi *et al.*, 2007). In contrast, high level resistant strains, VRSA, arise by acquisition of the *vanA* operon (carried on transposon Tn1546 on a conjugative plasmid) from vancomycin-resistant enterococci (VRE) (reviewed in Périchon and Courvalin, 2009) To date, 11 VRSA strains have been described, the first of which was isolated in the USA in 2002 (Zhu *et al.*, 2008). Of these 11 isolates, 9 have been characterised in the USA (7 in Michigan state), and the remainder in Iran and India (Zhu *et al.* 2008, Saha *et al.* 2008, Aligholi *et al.* 2008, Périchon and Courvalin 2009).

Interestingly, MRSA and VRE are frequently associated in the hospital environment. Several studies have reported co-colonisation and co-infection in of MRSA and VRE in patients (Sigurdardottir *et al.* 2006, Han *et al.* 2009, Milstone *et al.* 2008). However, Severin and colleagues (2004a) showed that in *S. aureus* isolates resistant to both methicillin and vancomycin, treatment with both antibiotics rendered these strains non-viable. This is because PBP2a (induced in the presence of methicillin) cannot cross-link

muropeptides in the cell wall that have the D-ala-D-lac group necessary for VRSA to prevent the action of vancomycin (Severin *et al.* 2004b). Nevertheless, although this indicates a short-term mechanism by which MRSA/VRSA infections can be treated, this can easily become complicated in the future by spontaneous mutations preventing this mechanism of control. *S. aureus* with reduced susceptibility to quinolones, tetracyclines, macrolides and aminoglycosides have also been documented (Hiramatsu *et al.*, 1997).

1.2.1.3 The Staphylococcal Chromosome Cassette *mec* (SCC*mec*) element

1.2.1.3.1 The SCC*mec* gene complex

mecA, a 2.1kb gene, is embedded into a larger 21-67kb mobile genetic element termed the *mec* element or staphylococcal chromosomal cassette (SCC*mec*). SCC*mec* is incorporated into the *S. aureus* chromosome at a site-specific location (*attBscc*) downstream of the *orfX* gene of unknown function, located near the origin of replication (Hiramatsu *et al.* 2001, Hiramatsu *et al.* 2002, Holden *et al.* 2004, Kuroda *et al.* 2001, Oliveira *et al.* 2002, Ito *et al.* 1999). It has been characterized as a mobile genetic element, separate from bacteriophage and transposons that renders the organism resistant to methicillin (Ito *et al.* 1999, Katayama *et al.* 2000, Hiramatsu 2001, Berger-Bachi and Rohrer 2002).

The SCC*mec* gene complex is essentially composed of three primary elements;

i. a *mec* element containing *mecA*, *mecR*, IS/*mecI*, IS*431*
ii. a *ccr* complex containing recombinase genes *ccrA/ccrB* or *ccrC*
iii. three joining regions (J regions; previously known as junkyard) consisting of 15-base direct and/or inverted core repeat sequences (Highuchi *et al.*, 2008).

Thus SCC*mec* can be summarised as: J3-*mec*-J2-*ccr*-J1 (Deurenberg and Stobberingh 2008, De Lencastre *et al.* 2007).

1.2.1.3.2 Classification of *SCCmec*

To date, five major classes (A-E) of *mec* complex have been characterised and are summarised in Table 1.3 (Taken from Deurenberg and Stobberingh ,2008).

Table 1.3 The major classes of *mec* complexes. Taken from Deurenberg and Stobberingh, 2008.

Class	Structure	SCC*mec*	Species
A	*mecI-mecR1-mecA-IS431*	II, III	*Staphylococcus aureus*
B	*IS1272-ΔmecR1-mecA-IS431*	I, IV, VI	*S. aureus*
C	*IS431-ΔmecR1-mecA-IS431*	V, VII	*S. aureus*
D	*ΔmecR1-mecA-IS431*	–	*Staphylococcus caprae*
E	*ΔmecR1-mecA-IS431*[a]	–	*S. aureus*

[a] 976 bp deletion in *mecR1* compared to class D *mec* complex.

The *ccr* gene complex encodes recombinases of the invertase/resolvase class and are located on all SCC*mec* elements. By mediating the integration of the SCC*mec* element from the recipient chromosome at the SCC*mec* attachment site (*attBScc*), these enzymes control the mobility of the SCC*mec* element (Zhang *et al.,* 2009). Three *ccr* genes (*ccrA*, *ccrB* and *ccrC*) have been documented, of which five allotypes have been described; four allotypes for the *ccrAB* genes (1, 2, 3, 4) and one for the *ccrC* gene. It is the different combinations of the *mec* and *ccr* complexes that comprise various SCC*mec* types. Further sub-types are classified according to variations in the J region DNA; although these are non-essential components of SCC*mec* element, additional antimicrobial resistance determinants (and genes encoding heavy metal resistance) may be carried in this region (Ito *et al.,* 2003). The drug resistance genes encoded by SCC*mec* elements are listed in

Table 1.4 (Ito *et al.* 2001, Ito *et al.*2003, Leclercq 2002, Oliveira *et al.* 2006, Deurenberg and Stobberingh 2008).

Table 1.4 Non β-lactamase resistance genes encoded by elements integrated into SCC*mec*

Element	Encoded gene	Confers resistance to
Plasmid pUB110	*ant (4')*	Aminoglycosides e.g. kanamycin, tobramycin, bleomycin
Plasmid pT258		Penicillins and heavy metals (e.g. mercury)
Plasmid pT181		Tetracycline
Transposon Tn554	*ermA*	Macrolide, lincosamide and streprogramin (MLS) resistance
ψTn554		Cadmium

Presently, eight SCC*mec* subtypes and variants have been described (Zhang *et al.* 2009, Ito *et al.* 2004, Derensinsk 2005, Shore *et al.* 2005, Qi *et al.* 2005, Jansen *et al.* 2006, Milheirico *et al.* 2007, Heusser *et al.* 2007, Kondo *et al.* 2007). In the original classification, these subtypes were designated roman numerals to determine type and were thus designated I-VIII. Figure 1.2 provides a schematic representation of SCC*mec* types I-VII (taken from Deurenberg and Stobberingh, 2008). SCC*mec* type IV is the most variable type, for which 8 subtypes (IVa-h) have been described, differing predominantly in the J1 region (Ma *et al.* 2002, Ito *et al.* 2003, Kwon *et al.* 2005, Shore *et al.* 2005). As the smallest of the SCC*mec* types, its enhanced mobility could be the cause for its variability (De Lencastre *et al.* 2007). Recently, the SCC*mec* III element initially described as 66.9kb in size, has been shown to be a composite of two SCC*mec* elements, SCC*mec* III (35.2kb) and SCC*mercury* (31.8) harbouring *ccr*C, pI*1258* and Tn*554* (Chongtrakool *et al.*, 2006).

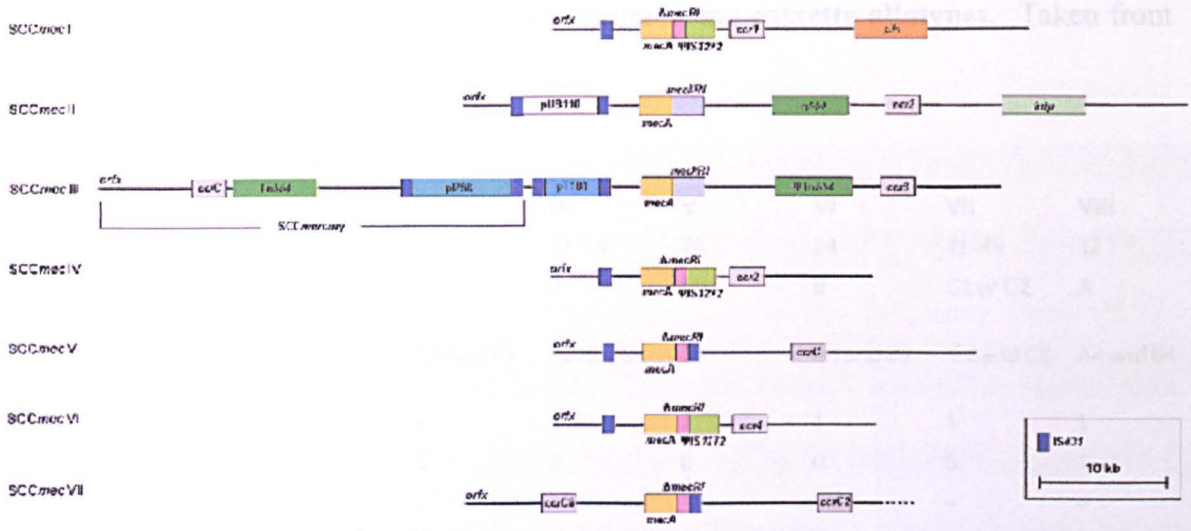


Figure 1.2 A schematic drawing of SCC*mec* types I to VII in MRSA. The major elements of the seven main SCC*mec* types (*ccr* genes, *IS 431, IS 1272, mecA, mecI /R1, orfX, pI258, pT181, pUB101* and *Tn554*) are presented (taken from Deurenberg and Stobberingh, 2008).

In comparison to the larger types (II and III), the smaller types (I, IV-VII) encode only β-lactam resistance (Deresinski, 2005). SCC*mec* types II and III also contain multiple determinants integrated into SCC*mec* for resistance to non-β lactam antibiotics and have been found to be responsible for the multidrug resistance commonly found in nosocomial MRSAs (Hiramatsu *et al.* 2001, Oliveira *et al.* 2002, Zetola *et al.* 2005).

Recently, a new nomenclature has been proposed for SCC*mec* types; defined by the combination of *mec* gene class and *ccr* allotype (Chongtrakool *et al.*, 2006). Based on this, the eight recognised SCC*mec* types are as follows (proposed names in parentheses): Type I (combination of the type1 *ccr* and the class B *mec* gene complex; 1B); type II (2A), III (3A), IV (2B), V (5C2), VI (4B), VII (5C1) and the latest, type VIII (4A). Table 1.5 compares the SCC*mec* allotypes (taken from Chambers and DeLeo, 2009).

Table 1.5 Comparison of staphylococcal chromosome cassette allotypes. Taken from Chambers and DeLeo, 2009.

Feature*	SCC*mec* allotype							
	I	II	III	IV	V	VI	VII	VIII
Size (kb)	34	53	67	21–24	28	24	41–49	32
mec complex	B	A	A	B	C2	B	C1 or C2	A
ccr complex	A1 and B1	A2 and B2	A3 and B3	A2 and B2	C	A4 and B4	C2 and C8	A4 and B4
IS*431* (n)	1	2	4	1	2	1	1	1
Tn*554* (n)	0	1	2	0	0	0	0	1
pUB110	–	+	–	–	–	–	–	–
pT181	–	–	+	–	–	–	–	–
pI258	–	–	+	–	–	–	–	–
Other resistance genes	None	*erm*, *spc* and *tobra*	*erm*, *tet* and Hg++	None	None	None	None	*erm* and *spc*

*mec complex A has intact regulatory genes, *mecR1* and *mecI*, upstream of *mecA*; *mec* complex B has regulatory gene deletions resulting from the insertion sequence 1272 (IS*1272*) insertion; *mec* complexes C1 and C2 have regulatory gene deletions resulting from the IS*431* insertion; the *ccr* complex is the recombinase locus; pUB110, pT181 and pI258 are plasmids integrated at insertion sequences. *erm*, erythromycin resistance gene; Hg++, mercury resistance gene; IS*431*, insertion sequence 431; n, number of copies; *spc*, spectinomycin resistance gene; *tet*, tetracycline resistance gene; Tn*554*, transposon 554; *tobra*, tobramycin resistance gene.

1.2.1.3.3 Origins of the SCC*mec* element

Analysis of other staphylococcal species has shown that *mecA* is not native to *S. aureus*, and must therefore have been acquired horizontally, probably from methicillin-resistant coagulase-negative staphylococci (MRCNS) (Wielders *et al.* 2001, Berger-Bachi and Rohrer 2002, Katayama *et al.* 2003, Wisplinghoff *et al.* 2003, Hansenn *et al.* 2004). Investigations of methicillin-resistant *S. epidermidis* (MRSE) and other MRCNS have detected the presence of SCC*mec* types I-V, as well as novel SCC*mec* elements (Hanssen *et al.* 2004, Hanssen and Sollid 2007, Wisplinghoff *et al.* 2003, Miragaia *et al.* 2007). It has been suggested that these novel SCC*mec* elements in MRCNs could act as an SCC*mec* pool for *S. aureus* (Corkill *et al.* 2004, Mongkolrattanothai *et al.* 2004). Zhang and colleagues (2009) in their recent classification of SCC*mec* type VIII, suggest this unique element could be the product of homologous recombination between two *S. epidermidis* strains, acquired by horizontal transfer into *S. aureus*. Other studies suggest *mecA*

descended from a close homologue ubiquitous in both β-lactam-susceptible and –resistant isolates of the animal commensal species *S. sciuri* (Antignac and Tomasz 2009, Couto *et al.* 1996, Couto *et al.* 2000, Wu *et al.* 2001, Oliveira *et al.* 2002). The *pdpD* gene, encoding PBP4, has been shown to be highly structurally similar to the *mecA* gene. In a recent study, Antignac and Tomasz (2009) showed that replacement of SCC*mec* from MSSA strain COL-S (an MRSA strain from which SCC*mec* was excised) with the *S. sciuri pdpD* gene restored the typical MRSA phenotype of the original MRSA strain COL.

In contrast to the common theory that MRSA emerged (over a very short period) under the selective pressure of β-lactamase-resistant antibiotics, the complexity of the *mecA* resistance mechanism prompted the proposal of alternative theories. De Lencastre and colleagues (2007) postulate the evolution of *mecA* resistance mechanisms occurred over a long time span, in a penicillinase-free staphylococcal species, under the selective pressure of penicillin. They propose this species to be *S. sciuri*, which is free of a penicillinase plasmid. Penicillin was used extensively in veterinary medicine as a prophylactic agent in 1949, very soon after its introduction and use in clinical medicine. Prophylactic use of penicillin could have been the selective pressure promoting the emergence of the *S. sciuri mecA* homologue (De Lencastre *et al.*, 2007). Others have proposed an element in *Maccrococcus caseolyticus* (formally *Staphylococcus caseolyticus*; Schleifer *et al.* 1982) to be the ancestral precursor to SCC*mec* (Baba *et al.*, 2009). Additional studies are necessary to determine the source of this genetic element.

1.2.1.3.4 Non-*mec* SCC elements

The SCC element, defined as the "*mobile chromosomal cassette with dedicated recombinase genes (ccr) and characteristic flanking short sequence repeats*" (De Lencastre *et al.* 2007) is not limited to dissemination of *mecA*. Several non-*mec* SCC and ψSCC (lacking *ccr* genes, or possessing non-functional ones) carrying other genetic elements have been documented. Genes encoded on these elements could contribute to the fitness/pathogenic potential of the genus. These are summarised in Table 1.6 (data from De Lencastre *et al.*, 2007). Several of these non-*mec* SCC and ψSCC have been found in coagulase negative staphylococci (CNS), further supporting for the importance of this group in the dissemination of SCC elements.

Recent whole genome sequencing of *S. haemolyticus* strain JCSC1435, and its comparative analysis with *S. aureus* and *S. epidermidis* species highlighted a chromosomal region downstream of the origin of replication (designated the "oriC environ") as important in the diversification and evolution of these clinically significant staphylococcal species (Takeuchi *et al.*, 2005). The analysis showed that although the oriC environ did not harbour genes essential for bacterial viability, species-specific genes were evident. The authors propose that SCCs were the driving vehicle for the introduction of exogenous genes into the oriC environ. Furthermore, it is postulated that other recombinases and IS elements may have served to remove non-beneficial genes, thus leaving only those species-specific genes which survived selection.

Table 1.6 Non-*mec* SCC elements.

Some of these elements encode putative virulence factors that may contribute to the fitness- or pathogenic-potential of the bacterium (information from De Lencastre *et al.*, 2007).

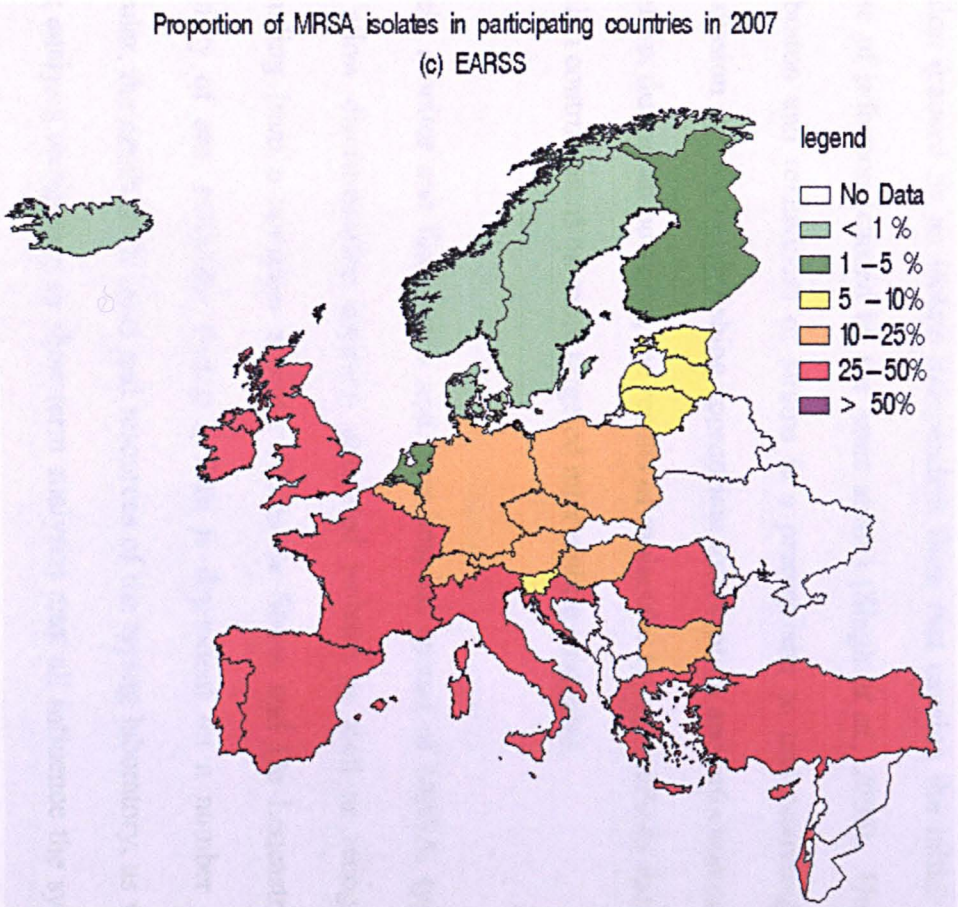
SCC element	Function	Reference
SCC*mer*	Heavy metal resistance	Chongtrakool *et al.*, 2006
SCC *C1*	Heavy metal resistance; DNA protection by restriction modification systems	Mongkolrattanothai *et al.*, 2004
SCC *MSSA476*	Fusidic acid resistance	Holden *et al.*, 2004
SCC *cap1*	Capsule biosynthesis	Luong *et al.*, 2002
SCC *15305cap*	Capsule biosynthesis	Kuroda *et al.*, 2005
ψSCC *h1*	Potassium transport	Takeuchi *et al.*, 2005
SCC*pbp4*	Cell wall cross linking	Mongkolrattanothai *et al.*, 2004
ψSCC *ACME*	Arginine deaminase and oligopeptide permease	Diep *et al.*, 2006

1.2.1.4 Epidemiological typing of *S. aureus*

1.2.1.4.1 Molecular epidemiology

MRSA remain among the most challenging pathogens, responsible for a huge burden of disease in healthcare facilities worldwide. The global spread of MRSA seems to be primarily due to the dissemination of a few pandemic clones in which the SCC*mec* element was acquired by a genetic background with potential for epidemic spread (Oliveira *et al.*, 2002). The prevalence of MRSA varies among different countries throughout Europe. This is monitored by a European national surveillance network (EARSS; European antimicrobial resistance surveillance system) coordinated by the Dutch National Institute of Public Health and the Environment (RIVM). Surveillance data show the lowest incidence of methicillin resistance (<1%) among *S. aureus* isolates is reported by the Nordic countries, whilst Southern European countries report high incidences of MRSA of

between 30—40%. Figure 1.3 illustrates MRSA prevalence data reported for 2007 and 2008.



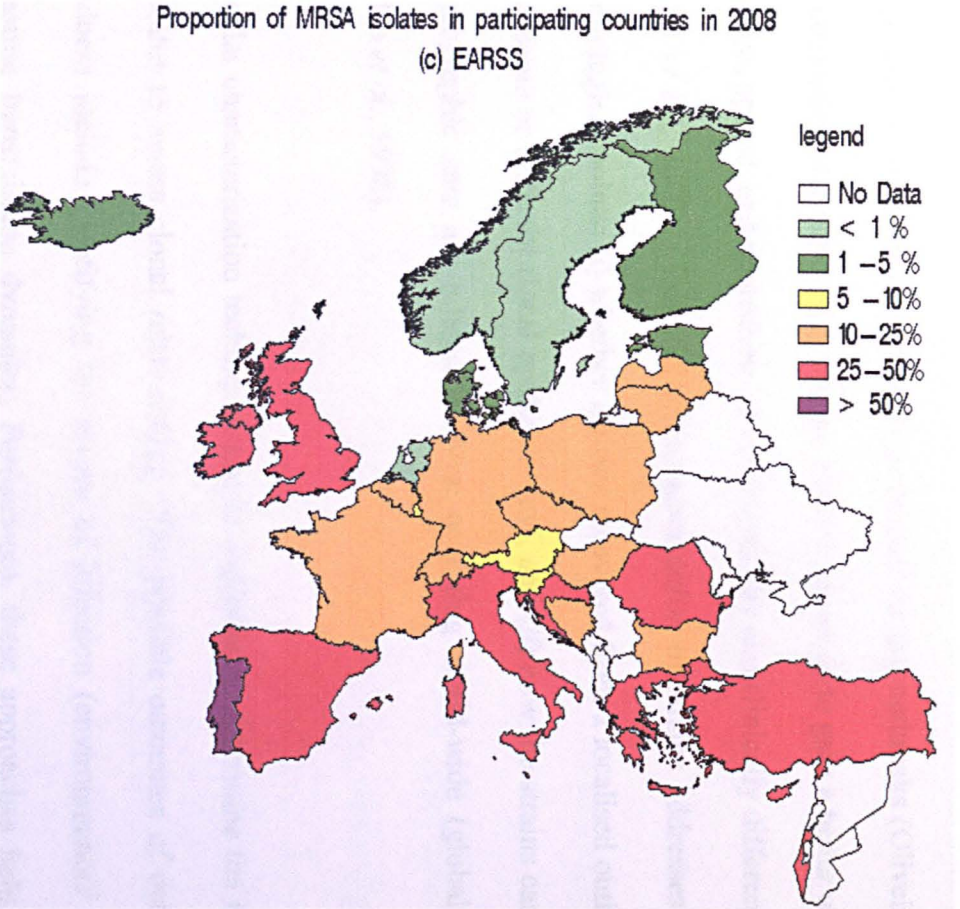


Figure 1.3 MRSA prevalence in Europe (EARSS data) reported in 2007 and 2008.

Molecular-based techniques greatly improve the capacity of hospitals to track the source and transmission dynamics of bacterial pathogens during outbreaks (Oliveira *et al.*, 2002). Characterisation of isolates of *S. aureus* enables scientists to gain a better understanding of the origins, spread, and relatedness of geographically and clinically different MRSA clones (Enright *et al.*, 2000). Essentially, characterisation of strains addresses two important epidemiological points: (*i*) whether isolates recovered from a localised outbreak of disease are the same or different (local epidemiology), and (*ii*) how the strains causing disease in one geographic area are related to those occurring world-wide (global epidemiology) (Maiden *et al.*, 1998).

Molecular characterization techniques can be exploited to determine the inter-relatedness of isolates to assess clonal relationships. The possible outcomes of determining strain relatedness include identifying the source of infection (environmental/ personnel) and delineating transmission dynamics. Furthermore, these approaches help distinguish re-infection (caused by an isolate independent from that causing the initial infection) from relapse of infection (caused by the same strain) (Singh *et al., 2006*). Understanding the distribution and relatedness of strains is a prerequisite to understanding the pattern of transmission and thus establishing appropriate intervention and infection control measures. Therefore, during an outbreak, the results of molecular typing methods focus the effects of infection control teams on more targeted intervention strategies.

To help monitor and limit intra- and inter-hospital spread of MRSA, typing techniques must allow discrimination between unrelated isolates as well as recognition of those descending from a common ancestor (Aires de Sousa and De Lencastre, 2004). The suitability of any particular typing system is dependent on a number of factors. In particular, the needs, skill level and resources of the typing laboratory, as well as the goal of the analysis (long-term or short-term analysis) may all influence the system used. An

ideal typing system needs to be capable of high typeability, adequate stability, display high technical reproducibility and high discriminatory power. Furthermore, it should be rapid, easy to perform and interpret, accessible, and cost-effective (Struelens, 1996). Key criteria can be defined as follows:

- Typeability requires the technique to unambiguously assign a result (type) to each isolate.
- Reproducibility refers to yielding the same result for an isolate upon repeat testing.
- Discriminatory power reflects ability to discriminate epidemiologically unrelated isolates, ideally assigning each isolate to a type.

1.2.1.4.2 Laboratory procedures for epidemiological analysis

Traditional phenotypic characterization methods for *S. aureus* include biotyping, serotyping, phage typing and antimicrobial susceptibility testing. However, genotypic identification methods (based on molecular analysis) have become more established over the last two decades; these provide in-depth information regarding strain relatedness (Singh *et al.* 2006). Techniques include pulsed-field gel electrophoresis (PFGE), other restriction-based methods, plasmid analysis, PCR-based typing methods and microarrays.

1.2.1.4.2.1 Phenotypic characterization methods for *S. aureus*

Biotyping (*bio*chemical *typing*), a species (and sub-species) identification method, can be used to differentiate strains based on properties such as differences in biochemical reactions and morphological appearance. Biotyping systems (e.g. API and Vitek) have now become automated and are used in some routine laboratories for species identification. However, strain resolution of *S. aureus* and coagulase-negative staphylococci is sub-

optimal, with poor reproducibility and weak discriminatory power (Low, 1997). As many of the phenotypic methods depend on the expression of metabolic or morphological features which may vary between strains, reproducibility may be limited (Becker *et al.*, 2004). Therefore, with respect to comparisons of *S. aureus*, such methods may not be ideal (Aires de Sousa and De Lencastre, 2004).

Antimicrobial susceptibility testing (e.g. broth/ agar dilution or disk diffusion methods) can be used to determine the resistance or susceptibility of an organism to a panel of antimicrobial agents. Although these methods are generally standardised and reproducible within and between laboratories, antibiogram data have limited value for epidemiological investigations because of their limited discriminatory power i.e. clonally unrelated isolates can have the same susceptibility patterns. Furthermore, the loss or acquisition of mobile genetic elements may confound results; the instability of these elements can adversely affect the interpretation of data. Genotypic methods can be used to monitor antimicrobial resistance genotypes as an alternative to (or to augment) phenotypic testing.

Serotyping of *S. aureus* is based on the detection of capsular polysaccharide proteins. Several serotypes of capsular polysaccharides have been reported to date, however, only 2 serotypes (types 5 and 8) are expressed by the vast majority (90%) of *S. aureus* clinical isolates. Capsular typing therefore has very poor discriminatory power as a strain typing system.

Phage typing has been used to distinguish bacteria according to their pattern of susceptibility/ resistance profiles to a defined set of bacteriophage (Holmberg *et al.* 1984, Lina *et al.* 1993, Schlichting *et al.* 1993). Generally, isolates of the same strain are lysed by the same phage(s). Consequently, phage typing is useful for strain identification and differentiation. Whilst the technique is fairly rapid, cost-effective and lends itself to high-

throughput, it is difficult to set up and maintain. In addition, reliable interpretation is subjective and non-typability of some isolates can limit its utility.

Overall for *S. aureus,* the discriminatory power and strain typeability properties of the phenotypic methods tend to be lower than can be achieved using genotypic methods (Table 1.7 compares the characteristics of the phenotypic and genotypic typing systems). Where the control of costs has a high priority, the introduction of new molecular methods may be difficult. However, costly material and equipment for molecular identification may be balanced by labour and other costs associated with traditional methods. Savings may also accrue over time as molecular identification protocols are often applicable to a variety of species while phenotypic typing schemes are restricted to one or a few species (Singh *et al., 2*006).

Table 1.7 Characteristics of phenotypic and genotypic typing systems. Taken from Arbeit, 1997.

Typing System	Proportion Of Strains Typeable	Reproducibility	Discriminatory Power	Ease Of Interpretation	Ease Of Performance	Extent Of Use
Biotyping	All	Poor	Poor	Good	Excellent	Moderate
Antimicrobial susceptibility testing	All	Fair	Poor	Excellent	Excellent	Widespread
Serotyping	Most	Good	Poor	Good	Fair	Rare
Phage typing	Most	Fair	Fair	Fair	Poor	Limited
Plasmid analysis	Most	Good	Fair	Good	Excellent	Moderate
Ribotyping	All	Excellent	Fair	Good	Good	Moderate
PFGE	All	Excellent	Excellent	Excellent	Good	Widespread
PCR analysis	All	Excellent	Good	Excellent	Good	Limited
Nucleotide sequence analysis	All	Excellent	Excellent	Excellent	Fair	Limited

1.2.1.4.2.2 Genotypic identification techniques

1.2.1.4.2.2.1 Genotypic characterization

The advent of molecular-based typing methods has overcome the shortcomings of phenotypic typing methods, in particular, improving strain typeability, standardization and reproducibility. The increased number of sequenced microbial genomes has enabled the development of several molecular typing systems to target single or multiple chromosomal loci. A principle of all molecular detection methods is that isolates with greater numbers of genetic differences are less likely to be epidemiologically related. Those molecular methods used for *S. aureus* typing can be characterised as follows (Aires de Sousa and De Lencastre, 2004):

- Plasmid analysis
- Pulsed-field gel electrophoresis (PFGE)
- Southern hybridisation analysis of digested DNA e.g. ribotyping
- PCR-based methods e.g. amplified fragment length polymorphism (AFLP) and SCC*mec* typing
- DNA sequence typing techniques e.g. *spa* typing and multilocus sequence typing (MLST)

Plasmid analysis

Plasmid analysis was the first molecular typing method used for bacterial analysis and the first DNA based typing system applied to *S. aureus* (McGowan *et al.* 1979, Archer *et al.* 1984, Meyers *et al.* 1976, Schaberg *et al.* 1981, Tenover 1985). Plasmids are self-replicating, mobile extra-chromosomal DNA elements found in the prokaryote cytoplasm.

Plasmid analysis as a typing tool involves isolation of plasmid DNA and comparison of the numbers and sizes by agarose gel electrophoresis. Where plasmids are large (e.g. 100-150kb), restriction digestion is incorporated into the method after plasmid isolation.

Plasmid analysis of *S. aureus* was shown to be useful only in typing MRSA (>90% carry plasmid), but not for typing methicillin sensitive *S. aureus* (MSSA) since approximately 50% lack plasmids, thus rendering some MSSA non-typeable (Coia *et al.*, 1988). Additionally, plasmid profiling has limited technical reproducibility due to features inherent to plasmid biology. Plasmids can exist in different molecular isoforms, for example supercoiled (closed circle), nicked (open circle), linear, and oligomeric. These isoforms will migrate at different rates during agarose gel electrophoresis, confounding the results and making the analysis difficult. Finally, epidemiologically unrelated isolates may have the same plasmid content whilst related isolates may have different plasmid content. Collectively, these problems demonstrate the limited reproducibility and discriminatory power of this system. Its usefulness is most probably effective only in the evaluation of isolates in a restricted time and place, e.g. during (suspected) acute outbreaks within a single hospital where it is expected that epidemiologically related isolates will have three or more plasmids in common (Mayer 1988, Schaberg and Zervos 1986). However, outside this scope, plasmid content evaluation has limited value in delineation of strain relatedness (Singh *et al, 2*006).

Pulsed-field gel electrophoresis (PFGE)

PFGE is generally accepted as the 'gold standard' for MRSA typing as it provides a high discriminatory index for microepidemiological (local outbreak) investigations. PFGE relies on the separation of restriction fragments of chromosomal DNA in a switching electric field to yield a banding pattern or 'fingerprint' following separation of bands on an

agarose gel. It provides a relative global overview of the chromosome, covering 90% of the genome (sum of the restriction fragment sizes). The banding pattern reflects the distance between infrequent restriction sites around the chromosome. Chromosomal variation (as occurs between unrelated isolates) will affect the restriction sites and/or the distances between them.

PFGE was first used for the analysis of *S. aureus* isolates in an intensive care unit (Prévost *et al., 1*991). It has since proven to be a very popular epidemiological tool with high discriminatory power. It is ideally used for studying local outbreak isolates but, although used, is less appropriate for large scale analysis or long-term epidemiology studies since small genetic changes are capable of radically changing PFGE banding fingerprints (Enright, 2008). In short, PFGE can identify small differences between isolates that are generally quite similar as occurs in a local outbreak (e.g. random genetic polymorphisms that may occur over the time of an outbreak; van Belkum *et al., 2*007). However, comparing large numbers of isolates is more problematic. In such a case, there is no reference from which to determine the degree of similarity that constitutes members of an outbreak strain that may change slightly over time. Clonal relationships may be difficult to define since minor sequence changes may result in significant pattern variation.

Although PFGE can be technically challenging, tedious and costly to perform, the biggest advantage is its ability to detect the movement of chromosomal mobile genetic elements. The greatest limitation of this technique is the difficulty of comparing complex fingerprints generated on different gels, especially when comparing between laboratories (Enright, 2008). Inter-laboratory standardisation methodologies in Europe (Murchan *et al. 2*004, Cookson *et al. 2*007) have only proved partially successful. The widespread use of PFGE has necessitated method standardisation, primarily to enable inter-laboratory comparisons (Murchan *et al. 2*003, van Belkum *et al.* 1998). Interpretation of fingerprints is not an

automated process and is therefore subject to variation according to the interpretation of the analyst locally. Guidelines have been proposed in an attempt to standardise the number of mismatches allowed in judging whether isolates are outbreak associated or non-outbreak associated (Tenover *et al.,* 1995). However, these guidelines are somewhat arbitrary, and therefore scientifically unsatisfactory.

Southern hybridisation analysis – Ribotyping

Some typing techniques are based on the use of restriction enzymes and electrophoresis to generate a DNA fingerprint (e.g. PFGE). Often the number of bands (DNA fragments) that can be resolved is limited. Southern hybridisation makes band resolution more manageable; chromatogram interpretation is simplified and conclusions about strain relatedness can be drawn more easily. During Southern blotting, the bands separated by electrophoresis are blotted (transferred) onto a nylon or nitrocellulose membrane. A labelled piece of DNA is then used to probe the membrane. Hybridisation of the probe to complementary sequences and subsequent probe detection enables the banding patterns to be resolved. As a typing method, the discriminatory power of this technique relies on the size and copy number as well as the positions of restriction sites on the targeted genomic sequence. The targeting of ribosomal RNA genes during Southern blotting is referred to as ribotyping.

Ribotyping has been used for species (and subspecies) identification as well as for strain typing. However, the discriminatory power of ribotyping has been shown to be lower than that of PFGE and some PCR-based methods (described below). Izard and colleagues (1992) described ribotyping as poor quality for differentiating coagulase-negative staphylococci. However, ribotyping can be highly automated, thus reducing human error

and limiting variation; the RiboPrinter microbial characterization system (Qualicon, Inc., Wilmington, DE) is an example of an automated ribotyping system (Singh *et al., 2*006).

PCR-based methods

PCR based methods rely on the amplification of targeted nucleic acid sequences so that only selected sequences are detected. It requires template DNA, complementary primers for each gene of interest and a heat stable DNA polymerase. Under optimal conditions, the primers bind to the target genes, serving as starting points for the polymerase to produce a complementary strand by the addition of nucleotide bases. Of those PCR-based methods used routinely for *S. aureus* typing, single-locus sequence typing (SLST) and multi- locus sequence typing (MLST, described further in the DNA sequence based typing section below) represent two of the most frequently used, targeting single and multiple genes respectively. Amplified fragment length polymorphism (AFLP) analysis based on linker-mediated amplification of selected restriction fragments is also used (Vos *et al.*,1995).

- AFLP

AFLP (Vos *et al.,* 1995) combines restriction digestion with PCR amplification. Several reports have documented the use of AFLP for typing *S. aureus* isolates (Grady *et al.* 2001, Cuteri *et al.* 2004, Melles *et al.* 2007, Savelkoul *et al.* 2007, Melles *et al.* 2009). Surgi and colleagues (2009) found AFLP to have a higher resolution in discriminating between MSSA populations compared with MRSA, due to the higher degree of clonality of the later. By combining AFLP with multiplex PCR, increased resolution of MRSAs was achieved. The technique is relatively easy to perform and the profiles are relatively easy to interpret for comparative analyses; software programs such as BioNumerics™ may be used to interpret the results. As a fingerprinting tool, it is comparable to PFGE, although

PFGE represents a greater portion of the genome sequence. AFLP is more reproducible than PFGE, and is therefore regarded to be more useful for epidemiological surveillance (Melles *et al.* 2007). However, as with PFGE, inter-laboratory reproducibility is sub-optimal, particularly when different electrophoresis platforms are employed (van Belkum *et al.*, 2007).

- SCC*mec* typing

Detection of the *mecA* gene (carried on the SCC*mec* element) in *S. aureus* has been recognised as the 'gold standard' for determining methicillin resistance (Murakami *et al.*, *1*991). With the increased knowledge of the variation that occurs within the SCC*mec* element (section 1.2.1.3), typing methods to detect the *mecA* gene as well classify the type of SCC*mec* element have been developed. Resistance genes for antibiotics and heavy metals present on insertion sequences, transposons and plasmids within SCC*mec* elements can also be detected (Enright, 2008). Compared with methods that only recognise *mecA* (which are not specific to just MRSA, but also coagulase-negative species that carry *mecA*), the SCC*mec* typing system represents a sensitive and rapid approach for obtaining results. Furthermore, *mecA* and SCC*mec* organisation are important epidemiological indicators of strain origin (Naimi *et al.* *2*003, Dufour *et al.* *2*002).

DNA sequence-based typing techniques

Sequence typing techniques can be divided into those that target a single locus (single-locus sequence typing; SLST) and those that target multiple loci (multi-locus sequence typing; MLST).

- SLST (*spa* typing)

The frequently employed SLST method for the analysis of *S. aureus* is *spa* typing; this involves the analysis of the polymorphic X or short sequence repeat region of the staphylococcal protein A gene (*spaA*). The polymorphic X region consists of 24-bp nucleotide repeat sequences that can vary in the number of repeats, due to deletions and multiplications of the units, as well as variations in the overall sequence due to point mutations (Koreen *et al., 2*004). This method, although less discriminatory than PFGE (Strommenger *et al. 2*006, Cookson *et al. 2*007), benefits from high throughput, ease of use and interpretation. Furthermore, it is valuable tool for multicentre studies since DNA sequence data can be controlled and checked, and the results curated digitally on online databases (www.spa.ridom.de) that can be queried remotely (Harmsen *et al. 2*003, Enright *et al. 2*008).

- MLST

MLST, based on the same principles as multi-locus enzyme electrophoresis (MLEE), was first developed in 1998 (Maiden *et al. 1*998) for the identification of hypervirulent lineages of *Neisseria meningitidis*. It subsequently became popular as a means of studying bacterial population structure of other pathogenic bacteria including *S. aureus* (Enright, 2000).

MLST involves amplification and sequencing of amplicons (450-500bp in length) of several genes (7 for *S. aureus*) encoding conserved metabolic compounds and comparing the results to an international database (www.mlst.net) (Francoise and Schrenzel, 2008). Compared with SLST, MLST probes a larger portion of the genome. The targets of MLST

are typically the internal regions of a series of housekeeping genes present in all isolates of a given species. Genetic polymorphism within these fragments is representative of a distinct allele. Isolates are thus defined by the alleles of all the target genes, which together comprise the allelic profile (sequence type) for that isolate. The basic principle of the typing profile is that individual allelic profiles do not occur by chance, and the potential alleles at each locus are many. Therefore, where isolates have the same allelic profile, they are considered as members of the same clone.

The target housekeeping genes are chosen to allow sufficient chance for evolutionary diversification in the absence of selective pressure. This technique is therefore well suited for epidemiological studies. However, as a method focused on conserved loci and not mobile genetic elements, MLST is limited in its discriminatory power for strain resolution. Additionally, the turnaround time may not be rapid enough for clinical surveillance for diagnostic purposes. MLST is relatively expensive and labour intensive compared with SLST due to the need to survey multiple genes. Sequence-based typing systems such as SLST and MLST are reproducible methods that are highly applicable to standardisation and database cataloguing, thus facilitating an internationally agreed nomenclature. As sequence data are unambiguous, quality control measures can easily be implemented to ensure accurate nucleotide assignment. Furthermore, sequence traces, allelic data and information about the organism can be collated over the internet to moderate the accuracy of information in the database (Enright, 2008).

1.2.1.4.2.2.2 Advanced molecular typing

1.2.1.4.2.2.2.1 Combining genotypic typing techniques

Based on MLST data the *S. aureus* New York/Japan clone and the Paediatric clone both belong to ST5. The combination of SCC*mec* analysis with MLST provides more useful information. Pairing these techniques can distinguish the New York/Japan clone (ST5-II) from the Paediatric clone (ST5-IV). In general, combining different typing techniques provides broader insights into the organism under investigation. *Spa* typing and PFGE are useful for fine strain typing e.g. microepidemiological investigations. MLST, *spa* and SCC*mec* typing can be used to relate isolates to information held on databases as well as published literature (Enright, 2008). PFGE, *spa* typing and MLST each display approximately comparable levels of resolution and congruency for clonal grouping (Cookson *et al.* 2007, Enright 2008). The approach taken by an individual laboratory is highly dependent on the cost and availability of resources.

1.2.1.4.2.2.2.2 Complete genome analysis

Many current genotyping methods only identify lineage without addressing the presence of individual genes or various mobile genetic elements that carry virulence and resistance genes. Thus none of the techniques discussed thus far extensively define the genes that constitute the organism(s) under investigation (Dunman *et al.*, 2004). Molecular methods such as sequencing and/or multi-strain microarray analysis are now available for completely determining the gene content of individual strains.

The extensive work undertaken in sequencing the genomes of *S. aureus* isolates has led to the publication of 14 complete genomes to date. These resources have given researchers an unprecedented glimpse into the genetics and biology of *S. aureus*. Such information can be used to better understand and identify the evolutionary potential of *S. aureus* (Lindsay and Holden, 2004). However, the cost of whole genome sequencing still remains

its biggest constraint. Additionally, genome sequencing is unsuitable for rapid diagnostics or epidemiological applications.

1.2.1.5 MRSA

1.2.1.5.1 The epidemiology of MRSA

Molecular epidemiology studies have shown the population structure of *S. aureus* to be clonal with some dominant lineages that have evolved independently. The population structure of *S. aureus* was first proposed several decades ago in studies by Rountree (Rountree and Freeman 1955, Rountree and Beard 1958) using bacteriophage typing. During the 1950s, a particular type of penicillin-resistant *S. aureus* caused severe disease in both the hospital environment as well as the community. This strain (first found in Australia and Canada) was lysed by 'phages 80 and 81 (Rountree and Beard, 1958), and became known as the 80/81 strain. This strain spread globally during the 1950s and acquired resistance to several antibiotics. Furthermore, its enhanced virulence was noted to be due to the production of an unusual leukocidin (Donahue and Baldwin, 1966), later identified as Panton-Valentine leukocidin (PVL) (Robinson *et al., 2*005). Strain 80/81 spread rapidly in the UK where it was the causative agent of more than 60% of staphylococcal infections in England and Wales (Anonymous, 1959) in 1957 compared to 13% in 1954 (Staphylococcal Reference Laboratory, Colindale, UK). Subsequent studies on staphylococcal population structure borne out of the analysis of nosocomial isolates have added further to our understanding of this organism.

One of the earliest explanations relating to the molecular evolution of MRSA proposed the single clone theory, in which was suggested all MRSA clones descended from a common

ancestor and that SCC*mec* was introduced only once into the *S. aureus* genome (Kreiswirth *et al.* 1993). However, subsequent studies favour the multi-clone theory, which suggests that SCC*mec* was introduced into various *S. aureus* lineages on several occasions (Enright *et al.* 2002, Qi *et al.* 2005, Fitzgerald *et al.* 2001, Gomes *et al.* 2006). Robinson and Enright (2003) in their analysis of an international panel of 147 diverse MRSA isolates, proposed that MRSA emerged at least 20 times via acquisition of SCC*mec*. They also observed that acquisition of SCC*mec* by *S. aureus* was a more common occurrence (by a factor of 4) than the replacement of an SCC*mec* element of one type with another type.

Retrospective studies of the first MRSA clone (ST8) using MLST revealed it to be closely related to 1950s isolates of methicillin-susceptible *S. aureus* (MSSA) that were prevalent in Denmark and possibly also seen elsewhere in Europe (Crisostomo *et al., 2001*); the only difference identified was *mecA* and *SCCmec* in the MRSA strain (Katayama *et al., 2000*). A pioneering study by Enright and colleagues in 2002 investigating the evolutionary history of MRSA using MLST and PCR analysis, collated these results using a novel algorithm, based upon repeat sequence type (BURST). The study analysed 912 globally distributed *S. aureus* isolates (both MSSA and MRSA). BURST grouped together similar allelic profiles ($\geq$ 5 allele commonality with at least one other isolate), and showed that five lineages (clonal complexes; CC) dominate for nosocomial MRSA (CC5, 8, 22, 30 and 45). This algorithm later became the publicly available online tool eBURST (Feil *et al., 2004*). eBURST is capable of predicting an ancestral progenitor (sequence type; ST) for each CC, determined by the genotype with the highest number of single-locus variants (SLVs), or in the case of a tie, double-locus variants (DLVs) (Enright *et al., 2008*). These (and similar) studies concluded that the MRSA clones stem from genetically similar MSSA that had acquired SCC*mec* elements, possibly on multiple occasions. In particular, Enright's study (2002) showed that ST8-MSSA (CC8) was the ancestor of the first MRSA

clone (ST250-MRSA-I); a single point mutation in *yqiL* locus in ST250 was the primary difference in the core genome. Schematic diagrams have been generated and illustrate the evolutionary descent of some of the major MRSA clones (fig. 1.4).

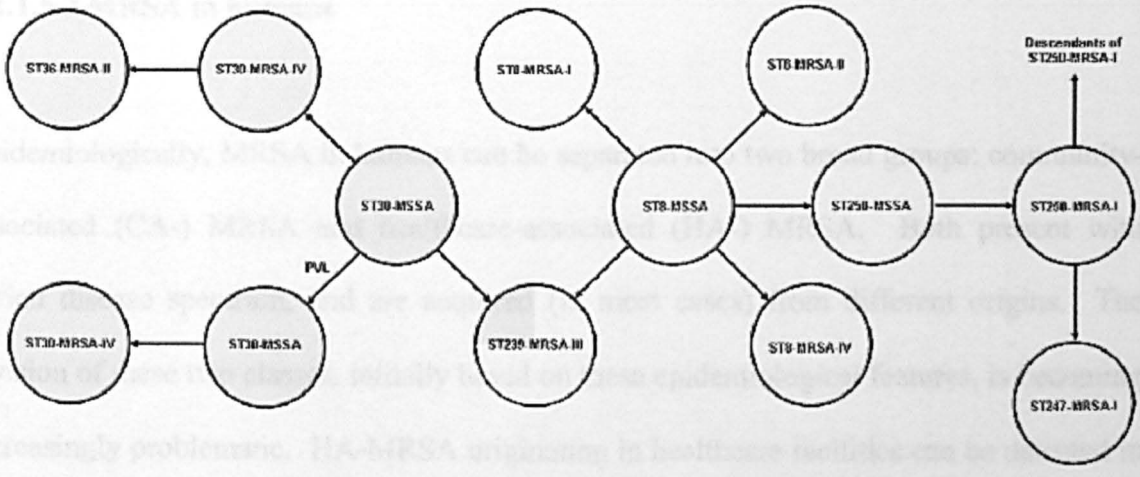


Figure 1.4 The evolutionary origins of the major MRSA clones and the possible relation between CA-MRSA and HA-MRSA. The arrows indicating either (1) the acquisition of SCC*mec*, (2) a change of SCC*mec*, (3) a change of ST, or (4) the acquisition of PVL. The grey coloured circles represent the MRSA clones from CC30, while the white circles represent the MRSA clones from CC8. ST239-MRSA-III from CC8 has evolved by the transfer of a 557-kb fragment from the chromosome of ST30 into a ST8 background.

Katayama and colleagues (2005) showed the dominant nosocomial MRSA lineages (CC1, 5, 8, 22, 30 and 45) were easier to transform with *mecA*-expressing plasmids compared with strains of other lineages. They therefore highlighted the fact that the genetic background of *S. aureus* as an important determinant for SCC*mec* stability. Furthermore, Noto and co-workers (2008) showed the genomic region around *attBscc* (site of integration of SCC*mec*) was variable, permitting SCC*mec* integration only in certain isolates. Collectively, these studies suggest the spread of a limited number of MRSA clones may be attributed to genetic backgrounds that favour integration and maintenance of SCC*mec*.

Further, Lindsay and co-workers (2006) have proposed the *S. aureus* restriction modification system determines the reception/rejection of SCC*mec* into these lineages.

1.2.1.5.2 MRSA in humans

Epidemiologically, MRSA in humans can be separated into two broad groups: community-associated (CA-) MRSA and healthcare-associated (HA-) MRSA. Both present with varied disease spectrum, and are acquired (in most cases) from different origins. The division of these two classes, initially based on these epidemiological features, is becoming increasingly problematic. HA-MRSA originating in healthcare facilities can be detected in the community, and thus appear to arise here (Tacconelli *et al.*, 2004). Similarly, CA-MRSA are no longer confined to the community as their transmission into the healthcare setting has been documented (HPA 2006, David *et al.* 2006, Otter and French 2006).

1.2.1.5.2.1 Healthcare-associated MRSA (HA-MRSA)

Acquisition of HA-MRSA is generally associated with recent hospitalisation, surgery, outpatient visit, nursing home admission, chronic illness, recent antibiotic exposure, injection drug use, and/or close contact with a person with risk factors for MRSA acquisition (Salgado *et al.* 2003, Aires de Sousa and De Lencastre 2004). With respect to the healthcare setting, the widespread use of broad-spectrum antibiotics and immunosuppressive drugs, and more adventurous surgery in an aging population are all factors that have favoured the rise in MRSA infections over recent years (Enright, 2008). Table 1.8 lists some of the internationally recognised HA-MRSA clones (discussed further in a recent review by Deurenberg and Stobberingh 2009).

Table 1.8 An overview of the major HA-MRSA clones. Taken from Deurenberg and Stobberingh, 2008.

HA-MRSA clone	MLST profile[a]	ST[b]	CC[c]	SCC*mec*[d]
Archaic	3-3-1-1-4-4-16	250	8	I
Berlin (USA600)	10-14-8-6-10-3-2	45	45	IV
Brazilian/Hungarian	2-3-1-1-4-4-3	239	8	III
Iberian	3-3-1-12-4-4-16	247	8	I
Irish-1	3-3-1-1-4-4-3	8	8	II
New York/Japan (USA100)	1-4-1-4-12-1-10	5	5	II
Pediatric (USA800)	1-4-1-4-12-1-10	5	5	IV
Southern Germany	1-4-1-4-12-24-29	228	5	I
UK EMRSA-2/-6 (USA500)	3-3-1-1-4-4-3	8	8	IV
UK EMRSA-3	1-4-1-4-12-1-10	5	5	I
UK EMRSA-15	7-6-1-5-8-8-6	22	22	IV
UK EMRSA-16 (USA200)	2-2-2-2-3-3-2	36	30	II

[a] Multilocus sequence typing.
[b] Sequence type.
[c] Clonal complex.
[d] Staphylococcal cassette chromosome *mec*.

Limiting MRSA spread within healthcare settings has proved challenging in many countries. Implementation of infection control policies and procedures are time-consuming and can be difficult to enforce. Some of the complications of limiting MRSA spread stem from the presence of unknown (hidden) reservoirs of carriers, as well as the emergence of highly epidemic clonotypes e.g. UK EMRSA-15 and EMRSA-16 (Auken *et al.* 2002, Cox *et al.* 1995, Richardson and Reith 1993, O'Neil *et al.* 2001). Screening high risk patients for MRSA colonisation has been suggested as a cost-effective measure for

limiting the spread of MRSA (Papia *et al.,* 1999). Early and reliable detection of MRSA carriage is crucial for infection control strategies. In this way, over prescription of last-line antibiotics (such as glycopeptides and oxazolidinones) can be prevented. The most recent Department of Health guidelines (published April 2009) requires all elective admissions (including surgical and medical day cases, except ophthalmic day cases) but not children, to be screened for MRSA. A decolonisation regime can then be prescribed for MRSA-positive patients prior to surgery. It is anticipated that this screening policy will be extended to all admissions, including emergencies, by 2010/11.

1.2.1.5.2.2 Community-associated MRSA (CA-MRSA)

1.2.1.5.2.2.1 Emergence of CA-MRSA

Until the 1990s, MRSA clones endemic within hospitals were rarely associated with healthy individuals who had no previous exposure to a healthcare environment. However, at that time, cases were reported in Australia and the United States (Okuma *et al., 2002*). CA-MRSA was first reported in Western Australia in 1993 from hospitalised patients who resided in remote communities (Udo *et al.,* 1993). Interest in the prevalence of CA-MRSA gathered pace following the 1999 report of four paediatric deaths in Minnesota and North Dakota, USA (CDC report 1999). Compared to nosocomial isolates, these MRSA had a relatively simple antibiotic susceptibility profile. They were resistant only to β-lactam antibiotics, possibly due to the carriage of SCC*mec* type IV which carries only the resistance gene to *mecA*. Furthermore, these isolates produced PVL toxin (Anonymous, 1999).

It is thought that MRSA colonisation within 'closed communities' such as the Australian aboriginals (Udo *et al. 1*993) and Native Americans (Groom *et al., 2*001) is likely to be

due to factors associated with spread in the community e.g. overcrowding and high rates of skin infections, as well as frequent use of broad-spectrum antibiotics (Maguire *et al.*, 1998). The earliest reports of serious CA-MRSA infections (particularly in the United States) occurred in populations of intravenous drug users, men who have sex with men, members of contact sports teams and prison inmates. However, increasing occurrences of infections in the general population have been reported (Moran *et al.*, 2006). Community-associated isolates have been associated with disease in younger healthier individuals with no previous risk associated with nosocomial disease (Chambers, 2001). These MRSAs often cause skin and soft-tissue infections, although cases of necrotising pneumonia are also reported with high fatality rates in young (median age 14 years), previously healthy individuals (Lina *et al.* 1999, Gillet *et al.* 2002). More recent studies have identified additional factors known as phenol-soluble modulins (PSM) which some workers suggest are important co-factors contributing to the enhanced virulence of CA-MRSA (Wang *et al.* 2007).

Many studies have focussed on CA-MRSA, however, no standard definition exists, with at least 8 different classifications available for describing MRSA infections as community-associated (Salgado *et al.* 2003). The Centre for Disease Control and Prevention (CDC) define CA-MRSA as strains isolated in an outpatient setting, or from patients within 48hrs of hospital admission. Other criteria include lack of a medical history of MRSA infection or colonization or history of recent (past year) of either hospitalisation (e.g. surgery), admission to a nursing home or dialysis, permanent indwelling devices e.g. catheters or other devices that pass through the skin (Deurenberg and Stobberingh 2009). Genetic markers such as SCC*mec* type, presence of the PVL genes and genetic background (lineage) have also been used to define CA-MRSA (Tristan *et al.*, 2007b).

1.2.1.5.2.2.2 Origins of CA-MRSA

Several groups have explored the origins of CA-MRSA. Some suggest SCC*mec* was acquired by *pvl*-positive MSSAs (which have a highly diverse genetic background) prevalent in the community (Ma *et al.* 2006, Taneike *et al.* 2006, Monecke *et al.* 2007). Conversely, Aires de Sousa and De Lencastre (2003) in their study on the evolution of HA-MRSA and CA-MRSA observed strong similarities between sporadic HA-MRSA isolates and CA-MRSA. They suggested that some MRSA described as community associated may have originated from hospitals. Robinson and colleagues (2005) propose the possibility of a common ancestor of HA-MRSA and CA-MRSA from their analyses of ST30 isolates; they propose ST30-MSSA as the ancestor. Phage type 80/81, a *pvl*-harbouring ST30 clone (penicillin resistant) emerged in the 1940s but disappeared following the introduction of penicillinase-resistant β-lactam antibiotics in the 1960s. It later re-emerged and acquired SCC*mec*IV to become the CA-MRSA SouthWest Pacific clone, ST30-MRSA-IV. Furthermore, it was also proposed that PVL-negative ST30-MSSA acquired type II SCC*mec*, possibly through transitional step including acquisition of SCC*mec* II, to become ST36-MRSA-II (the pandemic HA-MRSA, EMRSA-16). In contrast, other studies have shown CA-MRSA are clearly distinct from any of the major nosocomial MRSA clones (Okuma *et al.* 2002, Dufour *et al.* 2002), suggesting *de novo* evolution of CA-MRSA via horizontal acquisition of the *mecA* gene into different *S. aureus* genetic backgrounds (Slamenlinna *et al.*, 2002). Hallin and colleagues (2008) suggest the sporadic isolates may have originated in the community and subsequently moved into the hospital environment, but remained sporadic due to inept adaptation to this hostile environment. Vandenesch and colleagues (2003) showed the *pvl*-producing CA-MRSA carrying SCC*mec* IV belonged to six dominant lineages (ST1, 8, 30, 59, 80, and 93), which were (in most cases) distinct from those of nosocomial isolates (Table 1.9). In countries where international exchanges are frequent (e.g. Singapore), more diverse set of CA-MRSA have been documented (Deurenberg and Stobberingh 2009).

Table 1.9 An overview of the major PVL-positive CA-MRSA clones.

Taken from Deurenberg and Stobberingh, 2008.

CA-MRSA clone	MLST profile[a]	ST[b]	CC[c]	SCC*mec*[d]
European	1-3-1-14-11-51-10	80	80	IV
Southwest Pacific (USA1100)	2-2-2-2-6-3-2	30	30	IV
USA400	1-1-1-1-1-1-1	1	1	IV
USA300[g]	3-3-1-1-4-4-3	8	8	IV
USA1000	19-23-15-2-19-20-15	59	59	IV or VII

[a] Multilocus sequence typing.
[b] Sequence type.
[c] Clonal complex.
[d] Staphylococcal cassette chromosome *mec*.

1.2.1.5.2.2.3 Success and virulence of CA-MRSA

One of the most successful CA-MRSA to date is the so-called USA300 strain (*pvl*-positive ST8-IV, named according to its characteristic PFGE pulsotype), which was first reported in the USA in 2000. It spread rapidly across the USA causing outbreaks in Native American populations, military recruits, prison inmates, children in day care centres and among men who have sex with men. The appearance of USA300 (fig. 1.5) is thought to have arisen via the acquisition of SCC*mec* type IV by ST8-MSSA leading to the intermediate lineage USA500 (an ST8-SCC*mec* IV HA-MRSA strain), and subsequently led to the emergence of USA300 via acquisition of at least 20 genes, predominantly mobile genetic elements (Li *et al.* 2009, Tenover and Goering 2009, Enright *et al.* 2002, Deurenberg and Stobberingh 2008, Tenover *et al.* 2006).

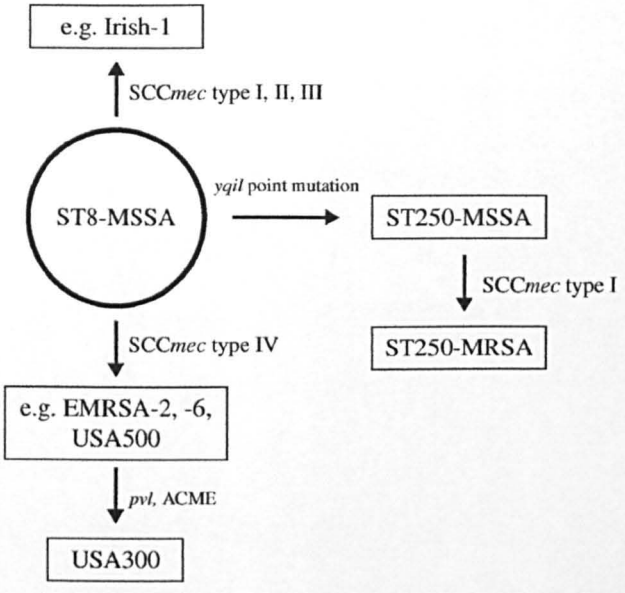


Figure 1.5 Evolutionary model of MRSA in CC8 leading to the emergence of USA300 and other pandemic clones of this lineage. Taken from Tenover and Goering, 2009.

At the time of the first report of USA300, another pulsotype USA100 (i.e. the New York/Japan clone, ST5-II) was the predominant clone recovered from HA infections. The rapid displacement of USA100 by USA300 led to intensive investigation and sequencing of the latter to determine factors implicated in its success (Diep *et al.,* 2006). Comparative genomics highlighted several unique mobile genetic elements carried by USA300, in particular *pvl* and the arginine catabolic mobile element (ACME) genes. In early reports, these mobile genetic elements were implicated in the increased transmissibility and virulence of this clone. ACME I found in USA300, is characterized by two gene clusters, *arc* and *opp-3*. The ACME-arc cluster (which differ from the native *arc* cluster found in all *S. aureus* isolates) encodes a complete arginine deaminase pathway involved at low pH and inhibition of the immune response against bacterial infections (Pi *et al.,* 2009). The ACME *opp-3* cluster, on the other hand, encodes a member of the ABC transporter family, with functions associated with peptide nutrient uptake, quorum sensing, pheromone transport, chemotaxis, eukaryotic cell adhesion, binding of serum components and expression of virulence determinants (Podbielski *et al.*, 1996). ACME is carried on a non-

mec SCC element known as SCC_{ACME}. More recently, Li and colleagues (2009) speculate core genomic virulence determinants (e.g. α toxin and phenol-soluble modulators which lyse human neutrophils) rather than accessory factors govern the virulence of USA300. This minor contribution of mobile genetic elements to the virulence of CA-MRSA is further supported by reports of CA-MRSA infections caused by *pvl*-negative strains (Diep and Otto 2008, Zhang *et al.* 2008).

Kennedy and colleagues (2008) conducted a study of the genomic sequences of 10 USA300 isolates recovered from across the USA from invasive MRSA infections spanning the clinical spectrum of diseases (uncomplicated bacteraemia to more invasive infections such as necrotising pneumonia, endocarditis, and osteomyelitis). They identified several single nucleotide polymorphisms (SNPs), of which the ratio of non-synonymous to synonymous mutations was 3:1. This study highlighted that the success of this clone could be due to its ability to acquire mutations (increasing its disease spectrum) rather than success due to the acquisition of additional virulence factors since genetic homogeneity (other than the SNPs) was evident in the isolates (Kennedy *et al.* 2008, Tenover and Goering 2009). Other studies support a key role for the accessory genome regulator, *agr* (the global regulator and quorum sensing element) in the enhanced virulence potential of USA300. Loughman and colleagues (2009) show that the USA300 clone harbours a highly active *agr* system. Montgomery and colleagues (2008) suggest this may have contributed to causing USA300 to displace USA400 as the predominant CA-MRSA strain in the USA. However, others refute the implication *agr* is responsible for differential virulence of USA300 and USA400 (Wang *et al.* 2007, Voyich *et al.* 2005). Tenover and Goering (2009) in their recent review of the origin and epidemiology of USA300 concluded that "*our mobile society, when coupled with the cramped and unsanitary living conditions of various populations, and the reluctance to culture wounds in emergency departments in the late 1990s and early 2000s probably led to the dramatic spread of this*

organism before it was recognised as a medical and public health problem". In summary, the success of the USA300 clone is attributable to a combination of several factors, including enhanced virulence, efficient colonization and host-to-host transmissibility (Li *et al.* 2009).

1.2.1.5.2.3 Classification: HA- versus CA-MRSA

The confusion with regard to classification lies in distinguishing true acquisition of MRSA infection in the community from MRSA detected in the community whose origin may be nosocomial. In recent years, CA-MRSA have started to displace HA-MRSA within healthcare institutions, particularly in areas of high prevalence of CA-MRSA e.g. USA/Taiwan (Otter and French 2006, Seybold *et al.* 2006, Klevens *et al.* 2006, Maree *et al.* 2007, Moran *et al.* 2006). An Australian study by O'Brien and colleagues (1999) reported introduction of an MRSA strain originating in the community into the hospital environment. However, the true origin of acquisition is difficult to determine since prior colonisation can persist from months to years, and the acquisition of MRSA may pass undetected unless clinical infection develops (Salgado *et al.* 2003). Therefore, one of the main criteria for distinguishing HA-MRSA from CA-MRSA lies in their phenotypic and genotypic characteristics. In contrast to HA-MRSA, CA-MRSA are (usually) susceptible to several non β-lactam antibiotics. MLST and PFGE analyses have highlighted differences in the genetic lineages of HA- and CA-MRSA (Groom *et al.* 2001, Naimi *et al.* 2001). CA-MRSA have been noted for their larger clonal diversity, suggesting that many *S. aureus* lineages have the potential to acquire smaller SCC*mec* elements and emerge as CA-MRSA (Enright *et al.* 2002, Okuma *et al.* 2002, Francois *et al.* 2008, Feng *et al.* 2008). Strain characteristics such as SCC*mec* type carriage, are also key to differentiating between HA-MRSA and CA-MRSA. The latter are primarily associated with the smaller SCC*mec* types (IV, V-VII). However, CA-MRSA associated with SCC*mec* types I, II or III have

been documented rarely (Cheung *et al.*, 2004). The high prevalence of *pvl* genes in CA-MRSA strains has also led to it being proposed as an epidemiological marker (Vandenesch *et al.*, 2003). A simplified summary of some of the main differences between these two groups is shown in Table 1.10.

Table 1.10 Differences between HA and CA-MRSA

HA-MRSA	Characteristics	CA-MRSA
Primarily healthcare (can overspill into community)	**Epidemiological setting**	Primarily community (some infiltration into healthcare)
Vulnerable patients particularly elderly or immunocompromised	**Individuals affected**	Young, previously healthy individuals
Wound infections to bacteraemia	**Clinical presentation**	SSTI to fatal necrotising pneumonia
Pandemic lineages	**Genetic background**	Diverse lineages
Various SCC*mec*	**Genetic traits**	SCC*mec* IV/V/VII
Multi-resistant	**Antibiotic resistance profile**	Resistant to β-lactams

SSTI, skin and soft tissue infections.

CA-MRSA have been shown to have a faster growth rate compared with HA-MRSA. As an important determinant of bacterial fitness, research into this area has recently led to the identification of the '*fudoh*' gene carried on SCC*mec*. Identified by a group in Japan (Kaito *et al.*, 2008), the *fudoh* gene (translation "spreading") was shown to suppress colony spreading, a property normally observed when *S. aureus* are spread on soft agar plates. Furthermore, *S. aureus* carrying the *fudoh* gene also displayed decreased exotoxin production and were less virulence in mice models. Kaito and colleagues showed *fudoh* to be associated with SCC*mec* II and III (typically found in HA-MRSA), the larger SCC*mec* elements, but missing from the smaller SCC*mec* IV and V elements. In strains lacking *fudoh*, high colony spreading ability was observed. Thus, this novel gene points to mobile genetic elements as possibly being the key to the success and international dissemination of MRSA carrying SCC*mec* types IV and V (predominantly CA-MRSA).

1.2.1.5.3 Livestock-associated MRSA: an emerging reservoir for CA-MRSA

Worldwide, there have been reports that livestock have become a reservoir of CA-MRSA. In the Netherlands (which typically report MRSA rates of <1%), a specific lineage of MRSA is shared by farm animals (in particular pigs and veal calves) and their handlers (Witte *et al,*. 2007). This was highlighted in 2003, when several unexpected cases of MRSA colonisation of humans in contact (direct/indirect) with pigs in the Netherlands were identified (Voss, 2005). Isolates of this particular lineage could not be typed by PFGE following *SmaI* digestion of genomic DNA, likely attributed to an unknown change in the DNA methylation system (Bens *et al.,* 2006). MLST analysis identified these isolates as belonging to sequence type ST398, an unusual lineage in humans. van Loo and colleagues (2007) in a retrospective case-control analysis, found a strong correlation between MRSA carriage and contact with either pigs or veal calves. Interestingly, although pig strains of ST398 are *pvl*-negative, documentation of *pvl*-positive human isolates highlight the ability of ST398 to acquire additional virulence determinants. Thus, unless adequately contained, ST398 has the potential to pose serious healthcare problems (Deurenberg and Stobbering, 2009).

The origin and emergence of this lineage within the animal population and subsequently in humans remains unclear. Some suggest ST398 was a highly prevalent MSSA strain in pigs which then acquired *mecA* from other staphylococci in pigs or their handlers (Wulf and Voss, 2008). Various SCC*mec* types have been found in the ST398 lineage, suggesting SCC*mec* may have been acquired on several occasions (van Duijkeren *et al.,* 2008).

The presence of an extensive network of pig farms within the relatively small country of the Netherlands could have facilitated the spread of ST398 (Wulf and Voss, 2008). This zoonotic MRSA has become a major problem in the Netherlands where a search-and-

destroy infection control strategy has been key to reducing MRSA rates. In particular, constant exposure of farmers to MRSA has made the 'destroy' part of the strategy (i.e. decolonisation) virtually impossible. The solution to the control of MRSA therefore lies in tackling the source of the problem, that is, MRSA in livestock. Revising policies for the use of antibiotics in livestock is a pressing issue. It is possible that the current animal husbandry regimes select for ST398 MRSA. With the potential to colonise and infect humans, further spread of this novel MRSA must be prevented; an effort involving clinicians, veterinarians and infection control practitioners (Wulf and Voss, 2008).

1.2.1.5.4 International surveillance of MRSA

The need for international surveillance systems with harmonized methodologies was recognised with the emergence and international spread of antibiotic resistance pathogens. One such system created in 1995 was the Centre for Molecular Epidemiology and International Network (CEM/NET) (De Lencastre, 2000). It was created to monitor and identify reservoirs of major multidrug-resistant *S. aureus* clones (and other Gram-positive pathogens). Several typing techniques were used for strain identification including PFGE, SCC*mec* typing, *spa* typing and MLST (Aires de Sousa and De Lencastre, 2004). The CEM/NET initiative analysed over 3000 MRSA isolates from Europe, Latin America, North America and Asia. Six lineages were identified as being particularly significant: Iberian, ST247-I; Brazilian, ST239-III; Hungarian, ST239-III; New York/Japan, ST5-II; Paediatric, ST5-VI and EMRSA-16, ST36-II. Such findings were recognised as evidence of the existence of only a few key pandemic clones which had spread worldwide (Oliveira *et al.*, 2002). Comparative genomics studies of staphylococcal genomes have led to several theories regarding the limited number of these pandemic clones. These have either supported intrinsic genetic determinants (i.e. the genetic background; De Lencastre *et al.* ,2007) or extrinsic environmental and host-associated determinants (Hallin *et al.*, 2008) as

factors promoting pandemic success. Some have suggested that epidemic lineages of MRSA are good colonisers, highly transmissible and virulent (Lipsitch 2001, Martinez and Baquero 2002). However, although these major MRSA clones have spread internationally, several reports have documented shifts in these MRSA clones over time. For example, Ma and colleagues (2006) showed that in Japanese hospitals between 1979-1985, the *pvl*-positive ST30-MRSA-IV clone was the dominant lineage, but was replaced by *pvl*-negative ST5-MRSA-II in the early 1990s.

In addition to the major epidemic clones, other lineages are also recognised; those that predominate in particular healthcare establishments but are not seen in others have been classed as minor clones. Furthermore, isolates recovered from an individual (or a few patients) in a single hospital have been termed sporadic isolates (Aires de Sousa and De Lencastre, 2004) to differentiate them from the highly successful lineages which have disseminated internationally.

In summary, rapid identification of MRSA in addition to knowledge about the evolutionary pathways (through studies of historic and extant collections) will greatly facilitate the overall aim of limiting the emergence and spread of MRSA. Furthermore, understanding MRCNS (implicated as a source for horizontal transfer of genetic elements in *S. aureus*) could provide insights into MRSA which may emerge in the future. This may be facilitated by microarray-based genome comparisons. With the readily available resource of sequence data on the internet (and as new isolates are sequenced), microarrays represent a powerful platform for analysing and comparing bacterial genotypes. Rather than comparing sequences base by base, as in traditional sequencing methods, genomes are compared gene by gene. This approach greatly reduces the cost of genome sequencing, whilst still providing data comparable with that of sequencing in terms of the presence or absence of genes (van Ijperen & Saunders, 2004). Gene arrays for bacterial genome

analysis have proven to be of great value in helping elucidate the genomic diversity and evolutionary relationship within species (Saunders *et al.*, 2004). Finally, elucidating the spectrum of virulence factors encoded by each isolate analysed will enable prediction of the disease potential of a particular strain (Li *et al.*, 2009).

1.3 REGULATION OF VIRULENCE

1.3.1 The staphylococcal virulence factor regulatory network

S. aureus strains possess genes encoding a wide variety of potential virulence factors, which are located either on the core chromosome or on mobile genetic elements. Virulence factors encompass bacterial surface proteins and secreted exoproteins whose biological activities include mediating host colonization, invasion of and damage to the skin and mucosa, dissemination through the body, and evasion of host defence mechanisms (Table 1.11). Other, potent factors have superantigenic properties (Projan *et al.* 1997, Dunman *et al.* 2001, Ferry *et al.* 2005).

The complex repertoire and regulatory network of *S. aureus* virulence factors is consistent with its ability to adapt to the human host in causing a wide array of infections (Bronner *et al.*, 2004). Research into the virulence mechanisms of *S. aureus* has concluded that with the exception of toxin-mediated diseases (e.g. toxic shock syndrome and staphylococcal scalded-skin syndrome), staphylococcal pathogenicity is generally not due to single virulence factors. Non-toxin mediated infections occur in a stepwise manner, involving a multiplicity of bacterial virulence and host factors (Jarraud *et al.* 2002, Ferry *et al.* 2005). The genes that encode the majority of these virulence factors belong to an extensive regulon that is co-ordinately regulated in response to a variety of both intra- and extracellular signals (Dunman *et al.*, 2001). Virulence factor production can be stimulated by changes in bacterial cell density, energy availability, environmental signals (pH, CO_2, O_2) and superantigens (Bronner *et al.*, 2004) and are carefully controlled so that they are only produced when required.

Table 1.11 Selected *S. aureus* virulence factors. Taken from Gordon and Lowy, 2008.

Type of virulence factors	Selected factors[a]	Genes	Associated clinical syndromes
Involved in attachment	MSCRAMMs (e.g., clumping factors, fibronectin-binding proteins, collagen, and bone sialoprotein-binding proteins)	*clfA, clfB, fnbA, fnbB, cna, sdr, bbp*	Endocarditis, osteomyelitis, septic arthritis, and prosthetic-device and catheter infections
Involved in persistence	Biofilm accumulation (e.g., polysaccharide intercellular adhesion), small-colony variants, and intracellular persistence	*ica* locus, *hemB* mutation	Relapsing infections, cystic fibrosis, and syndromes as described above for attachment
Involved in evading/ destroying host defenses	Leukocidins (e.g., PVL and γ-toxin), capsular polysaccharides (e.g., 5 and 8), protein A, CHIPS, Eap, and phenol-soluble modulins	*lukS-PV, lukF-PV, hlg, cap5* and 8 gene clusters, *spa, chp, eap, psm-α* gene cluster	Invasive skin infections and necrotizing pneumonia (CA-MRSA strains that cause these are often associated with PVL) abscesses (associated with capsular polysaccharides)
Involved in tissue invasion/penetration	Proteases, lipases, nucleases, hyaluronate lyase, phospholipase C, and metalloproteases (elastase)	*V8, hysA, hla, plc, sepA*	Tissue destruction and metastatic infections
Involved in toxin-mediated disease and/or sepsis	Enterotoxins, toxic shock syndrome toxin-1, exfoliative toxins A and B, α-toxin, peptidoglycan, and lipoteichoic acid	*sea-q* (no *sef*), *tstH, eta, etb, hla*	Food poisoning, toxic shock syndrome, scalded skin syndrome, bullous impetigo, and sepsis syndrome
With poorly defined role in virulence	Coagulase, ACME, and bacteriocin	*arc* cluster, *opp-3* cluster, *bsa*	

[a] Several factors may have >1 role in *S. aureus* pathogenesis. ACME, arginine catabolic mobile element; CA-MRSA, community-acquired methicillin-resistant *S. aureus*; CHIPS, chemotaxis inhibitory protein of staphylococci; Eap, extracellular adherence protein; MSCRAMMs, microbial surface components recognizing adhesive matrix molecules; PVL, Panton-Valentine leukocidin

In the laboratory setting, expression of many *S. aureus* virulence factors follows a predictable pattern. Cell wall associated adhesins (e.g. microbial surface components recognising adhesive matrix molecules, MSCRAMMs) are expressed during the early exponential growth phase when bacterial cell density is low. Conversely, haemolysins, toxins, and enzymes facilitating tissue destruction and dissemination (secreted extracellular proteins) are expressed at the end of the exponential phase and during the stationary phase (Projan *et al.*, 1997); figure 1.6.

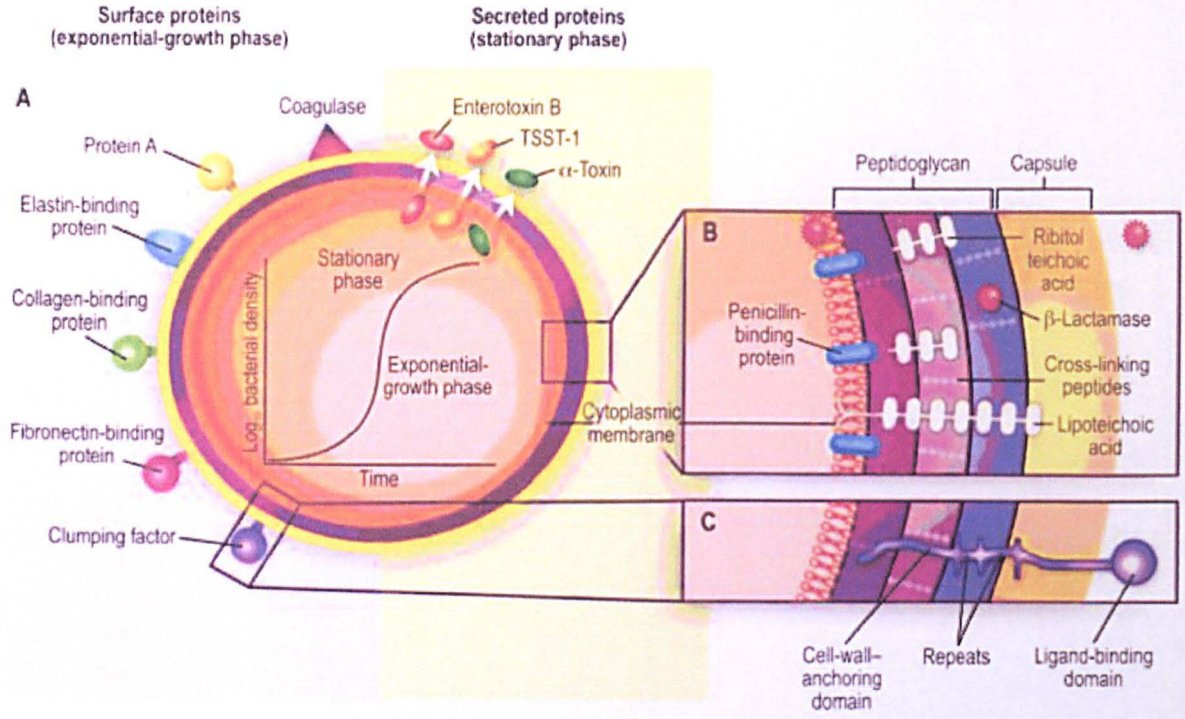


Figure 1.6 Growth phase dependant production of cellular components. Taken from Lowy 1998.

A. Growth-phase dependent synthesis of *S. aureus* virulence factors. Illustrated are the most recognized surface and secreted proteins.

B and C. Cross sections of the cell envelope. The structural organisations of many of the surface proteins are similar to that of clumping factor (TSST-1, toxic shock syndrome toxin).

In their review of the regulation of virulence in staphylococci Dunman and Projan (2001) posit that the purpose of virulence factors is to enhance the survival of the bacterium in adverse environments, rather than to cause disease. As such, the infected host represents an environment to which the bacterium must adapt for its survival and success. Few (if any) virulence factors have been shown to be essential for bacterial proliferation *in vitro*. However, many factors have been recognised as essential during infection (Dunman and Projan, 2001). Therefore, *in vitro* studies coupled with *in vivo* experimentation will be necessary to give more precise understanding of disease causation. Dunman and Projan also observed that cell wall-associated virulence factors require ongoing cell wall assembly to become embedded into the cell wall. Therefore, the exponential phase during which cell

division and proliferation occurs is the optimum time for production of these factors. Conversely, the effects of extracellular proteins are most pronounced when the bacterial cells are localised and in enough concentration to produce a "beneficial effect". Thus, the optimum time for production of extracellular proteins is during the post-exponential phase (Dunman and Projan, 2001).

Extensive research has revealed the significance of global regulators, such as the much studied Agr system (discussed further in section 1.3.2.1), in contributing to the regulatory network. The more research is conducted in this area, the more the complexity of this network has become apparent. Analysis of changes in gene expression, initially in culture, *in vitro*, and subsequently in animal models *in vivo*, has helped to delineate known and putative pathways of virulence regulation. This knowledge of the global regulatory network is continually being refined and reviewed (Novick *et al.* 1993, Founier *et al.* 2001, Dunman and Projan 2001, Baba *et al.* 2002, Garvis *et al.* 2002, Novick and Jiang 2003, Bronner *et al.* 2004, Ferry *et al.* 2005, Fournier 2008).

Several two-component regulatory systems that are sensitive to environmental signals have been identified within *S. aureus*. These systems generally consist of two proteins: a histidine kinase sensor and a response regulator. The sensor can be either associated with a receptor, or it can bind to an extracellular ligand. Upon binding of the extracellular ligand to the receptor, autophoshorylation occurs in which there is a transfer of a phosphate residue from ATP to a histidine residue of the cytoplasmic domain of the sensor kinase. This phosphate residue is then transferred to an aspartate residue on the response regulator. The end result of this cascade is the binding of the response regulator to specific DNA target sequences, permitting transcription-regulating functions to initiate. The effector molecules of the two-component systems can then be transcribed.

1.3.2 Global regulators of virulence gene expression

The main global regulators of *S. aureus* can be classed into 4 groups, the details of which are discussed below: (*i*) the Agr system, (*ii*) the SarA system and its homologues, (*iii*) the Sae system, and (*iv*) the Arl system.

1.3.2.1 The Agr system

1.3.2.1.1 Overview of the Agr system

The Agr locus was first discovered in the early 1980s when a mutant strain (carrying a Tn551 insertion at this locus) displaying a *hla*-negative phenotype was identified (Mallonee *et al.*, 1982). Later work suggested this locus probably encoded a regulator of the virulence genes *tsst*, *spa*, *hla*, *hlb* and *hld* (encoding toxic shock syndrome toxin, protein A, alpha toxin, beta toxin and delta toxin respectively) (Recsei *et al.*, 1986). Subsequently, Agr has been studied in extensive detail, and its mechanisms of action have been well characterized. Cloning and sequencing revealed a two-component regulatory system (Peng *et al.*,1988). However, further work highlighted the complexity of this system compared with that of classical two-component systems. The Agr locus is now known to consist of a four gene operon, *agrBDCA*. Transcription in the forward direction produces one RNA molecule, RNAII, whereas transcription in the opposite direction (overlapping the *hld* gene encoding delta toxin) produces another RNA molecule, RNAIII (Novick *et al.* 1993, Morfeldt *et al.* 1995). These two transcripts (initiated from two promoters, P2 and P3, respectively) are essential to *agr* mediated regulation. RNAIII, the main effector molecule of the Agr locus (Novick *et al.*, 1993), is produced in response to quorum sensing and functions to regulate mainly gene translation (although it is also known to act at the transcription level; Novick *et al.* 1993). Remarkably, it acts both as a

repressor and activator. It is positively regulated by the gene products of the *agrBDCA* operon, increasing its level during growth and reaching a maximum in the post-exponential phase (Vandenesch *et al.*, 1991).

As well as acting on other global regulators, *agr* activates the virulence factor *hla* (representing the classic secreted extracellular protein), but also suppresses *spa* (encoding protein A, the classic cell wall-associated protein) via RNAIII (Novick *et al.* 1993, Benito *et al.* 1998, Benito *et al.* 2000). This modulation of gene expression by RNAIII has been shown to operate based on conformational changes in the secondary structures of RNA molecules. It is the regulation of these two proteins that have been studied in detail with respect to the *agr* locus and its effects; it is for this reason that *hla* and *spa* are almost always depicted in the diagrammatical representation of the *S. aureus* virulence regulatory network .

The *agr* locus has been shown to be modulated via autocatalytic feedback (Novick *et al.*, 1995). AgrA and AgrC have been classified as the response regulator and histidine kinase components, respectively, in this complex two-component system (Novick *et al.* 1995, Lina *et al.* 1998). During the exponential growth phase, AgrD (a propeptide which is proteolytically digested and matured by AgrB) is secreted into the extracellular environment; the result is a mature autoinducing peptide (AIP). The concentration of AIP is linked directly to cell density. Once AIP concentration reaches a specific threshold, the Agr locus is able to regulate itself further; a process defined as quorum sensing. Binding of AIP to its receptor AgrC, causes AgrC to become phosphorylated and activates the response regulator AgrA by a second phosphorylation step. Activated AgrA is then able to stimulate transcription of the *agr* P2 and P3 promoters and thus the transcription of RNAIII, the major effector molecule of the *agr* response. Therefore, the AgrC protein is the receptor/sensor of AIP and AgrA is the response regulator in this signal transduction

pathway (fig. 1.7). The Agr system has two recognised functions effected via RNAIII, first it activates the transcription of several extracellular proteins and second it represses the transcription of several cell-wall associated proteins in a cell-density dependent manner (Janzon *et al.* 1986, Recsei *et al.* 1986, Gaskill and Khan 1988, Mahamood and Khan 1990, Cheung *et al.* 1992, Patel *et al.* 1992, Sheehan *et al.* 1992, Dassy *et al.* 1993, Daugherty and Low 1993, Tremain *et al.* 1993, Chamberlain and Imanoel 1996, Saravia-Otten *et al.* 1997, Arvidson 2000, Bronner *et al.* 2000, Novick 2000, Arvidson and Tegmark 2001, Dunman *et al.* 2001, Reed *et al.* 2001, Luong and Lee 2002, Schmidt *et al.* 2004, Xiong *et al.* 2004, Rooijakkers *et al.* 2006). Dunman and colleagues (2004) in their analysis of *agr*-controlled genes identified 104 upregulated and 34-downregulated genes. Furthermore, Zeibandt and colleagues (2004) showed similar up- and down-regulation of most of these virulence factors at the protein level. Interestingly, some studies have noted that *agr* mediated regulation of virulence factors depends on the strain used and also therefore on genetic background (Blevins *et al.* 2002, Cassatt *et al.* 2006).

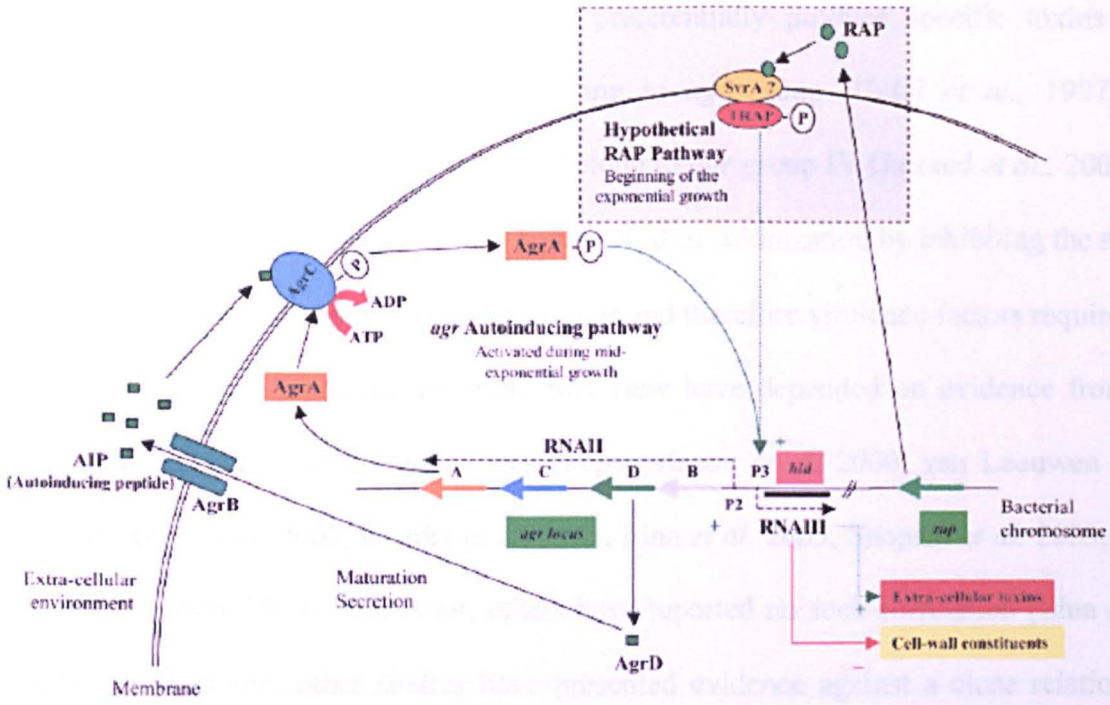


Figure 1.7 Schematic diagram illustrating autoregulation of the Agr locus, and its interactions with the TRAP (target of RNAIII activating protein) system (see text; taken from Bronner *et al.,* 2004).

Furthermore, strain variation in the Agr locus has been the basis for defining 3 allelic groups (*agr* types I, II and III). Ji and colleagues (1997) in their analysis of the Agr locus noted that in the laboratory strain RN6390B the Agr locus was either activated or inhibited by the culture supernatant of other strains. From this they deduced three groups of Staphylococci based on AIP cross-inhibition of the Agr response. Within groups, the AIP activated the Agr locus but AIP from outside the group inhibits the *agr* signal pathway. *S. aureus* AIPs also inhibit the *agr* pathways of other Staphylococcal species. Thus, each group is defined by inter-group inhibition of the *agr* response via AIP peptides (Ji *et al.,* 1997). A fourth *agr* group was later defined by Jarraud and co-workers (2000). Sequencing of Agr loci showed a high degree of conservation between strains of the same Agr group. The variation observed between groups was observed in the B-D-C region (Ji *et al.* 1997, Otto *et al.* 1999, Jarraud *et al.* 2000).

It has been noted that some *agr* groups preferentially produce specific toxins e.g. menstrual toxic shock strains mainly belong to *agr* group III (Ji *et al.*, 1997) and epidermolysin A producing strains mainly belong to *agr* group IV (Jarraud *et al.*, 2000). It has been suggested that *agr* group may be involved in colonization by inhibiting the spread and growth of other strains via inhibition of *agr* and therefore virulence factors required for spread and survival. Studies favouring this view have depended on evidence from the distribution of particular groups *in vivo* (Papakyriacou *et al.* 2000, van Leeuwen *et al.* 2000, Peacock *et al.* 2002, Goerke *et al.* 2003, Lina *et al.* 2003, Shopsin *et al.* 2003, Gilot and van Leeuwen 2004). However, others have reported no such correlation (Lina *et al.*, 2003). Furthermore, other studies have presented evidence against a close relationship between *agr* group and virulence (Peacock *et al.* 2002, Goerke *et al.* 2003, Kahl *et al.* 2003, Gilot and van Leeuwen 2004).

1.3.2.1.2 Activation of Agr via the RAP/TRAP system

An alternative mechanism for the activation of the Agr locus through an autoinducing peptide has been recognised and named the RAP/TRAP two component system. RNAIII-activating protein (RAP), the autoinducer, which is constitutively secreted (Balaban And Novick 1995) has been proposed to activate the phosphorylation of Target of RAP (TRAP), which binds to a membrane associated protein putatively identified as the staphylococcal virulence regulator (SvrA; Gillot *et al.* 2002, Balaban *et al.* 2001). Phosphorylated TRAP is able to activate *agr* P2/P3 promoters (Garvis *et al.*, 2002); see figure 1.7. Balaban and co-workers (2001) in their study of TRAP regulation suggested that AIP, produced during mid-exponential growth, inhibits TRAP phosphorylation by causing the activation of a phosphatase. They further suggested that RAP and AIP may both activate the Agr locus temporarily. That is, AIP (the mature, processed product of the *agrD* transcript) activates the Agr locus via phosphorylation of AgrC and activation of

AgrA, whilst RAP activates the Agr locus via phosphorylation of TRAP. Finally, an RNAIII-inhibiting peptide (RIP) has been recognised as an inhibitor of the Agr locus in coagulase negative *Staphylococcus* species *S. warnerii* and *S. xylosus* (Balaban *et al.*, 2001). It was proposed that RAP and RIP compete as agonist-antagonist in their binding for the same receptor for RNAIII activation.

1.3.2.1.3 Control of the Agr locus via other global regulators

Expression of the Agr locus has been shown to be under the control of other global regulators, e.g. staphylococcal accessory regulator (Sar; discussed below). Deletion of the *sarA* gene (in strains RN450 and RN6390) results in dramatically reduced synthesis of RNAII and RNAIII, demonstrating that SarA protein up-regulates *agr* expression (Cheung and Projan 1994, Morfeldt *et al.* 1996). Regulation occurs via binding of SarA to the P2-P3 promoter region (Heinrichs *et al.* 1996, Morfeldt *et al.* 1996, Cheung *et al.* 1997, Chien and Cheung 1998, Rechtin *et al.* 1999). Indeed at least four regulators (SarA, SarR, SarX and SarU) are known to bind to P2-P3, highlighting the complexity of the regulation of this locus (Manna and Cheung 2006a and 2006b). In addition, RNAIII transcription is under the indirect control of many other regulators. Examples include σ^B which down regulates RNAIII production (Bischoff *et al.* 2001, Horsburgh *et al.* 2002). RNAIII acts mainly by modulating target gene transcription. This can occur directly by binding to the promoters of target genes or indirectly by regulation of other transcriptional regulators e.g. SarT and SarS (Tegmark *et al.* 2000, McNamara *et al.* 2000, Schmidt *et al.* 2001).

1.3.2.2 The Sar system

The staphylococcal accessory regulator (Sar) locus generates three overlapping transcripts (*sarABC*) synthesised from three promoters (P1, 2 and 3 respectively), all of which

terminate at the same stem-loop sequence. All three transcripts of the Sar locus encode the message for SarA (fig. 1.8). SarA modifies expression of virulence genes either directly by binding to conserved DNA regions (Sar boxes) within promoter sequences e.g. *hla*, *spa*, *fnb, sec* or indirectly via suppression of RNAIII transcription by binding to the Agr P3 promoter (Novick *et al.* 1993, Saravia-Otten *et al.* 1997, Chien *et al.* 1999, Rechtin *et al.* 1999, Dunman *et al.* 2001).

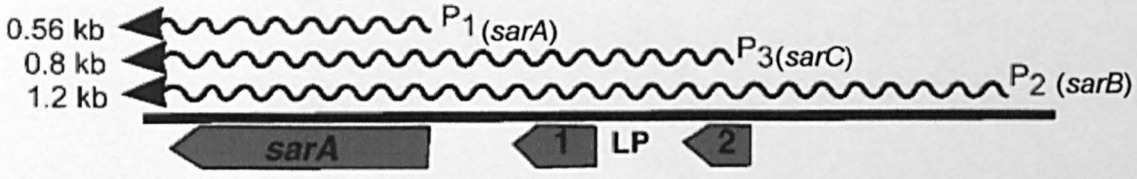


Figure 1.8 The SarA locus. The three transcripts, *sarABC*, are synthesised from the three promoters P1, 2, 3 respectively.

Unlike the Agr system (RNAIII) which up-regulates extracellular proteins and down-regulates cell wall-associated proteins during the post-exponential growth phase (Janzon *et al.* 1990, Kornblum *et al.* 1990, Morfeldt *et al.* 1995, Novick *et al.* 1993), SarA up-regulates both surface-bound and extracellular proteins (Cheung *et al.* 1992). The level of SarA protein produced during all growth phases remains constant, whereas the levels of SarB and SarC depends on growth phase (Blevins *et al.,* 1999). SarB is more prominent in the mid- to late-exponential phase whilst SarC is prevalent in the late-exponential to stationary phase (Bayer *et al.* 1996, Manna *et al.* 1998). Like Agr, the SarA locus is autoregulatory (Chien and Cheung 1998, Manna *et al.* 1998) since deletion of the *sarA* gene attenuates transcription of the other transcripts of this locus.

The regulation of SarA occurs by means of SarR, the regulatory protein that binds the *sarA* promoter (P_1). Expression of SarR peaks during the post-exponential phase and it decreases the transcription of *sarA* during the late exponential and stationary growth

phases (Manna *et al.*, 2001). SarA has been shown to be a strong regulator of protease genes such as serine protease (*sspA*), cystein protease (*scpA*) and aureolysin (*aur*), as well as fibronectin binding protein (*fnbp*) and protein A (*spa*) (Cheung *et al.* 1992, Lindsay and Foster 1999, Wolz *et al.* 2000, Karlsson *et al.* 2001, Karlsson and Arvidson 2002). Furthermore, SarA has been recognised as an important regulator of biofilm formation. It induces the intracellular adhesin (*ica*) operon (Beenken *et al.* 2003, Valle *et al.* 2003), as well as regulating *bap* expression, both of which are important in biofilm formation (Cucarella *et al.*, 2001). SarA is thought to regulate transcription by binding to DNA (Rechtin *et al.*, 1999), although the protein structure has not revealed a classic DNA-binding motif. Some have proposed that this regulation is mediated by conformational changes in DNA and DNA supercoiling (Schumacher *et al.* 2001, Roberts *et al.* 2006).

1.3.2.3 Sar Homologues

Sar homologues belong to the MarR winged helix family of transcriptional regulators (Ellison and Miller, 2006), comprising two subgroups based on protein size. Group-1 comprises the smaller size proteins (13-16kDa) which bind DNA as dimers; these include SarA, SarR, SarT and SarX. The second group comprising the larger proteins (29-30kDa) include Rot, SarS, SarT, SarU, SarV, MgrA and TcaR; these proteins also bind DNA but have N- and C-terminal regions similar to SarA (Liu *et al.* 2001, Schumacher *et al.* 2001, Li *et al.* 2003). Of these homologues, one of the most studied is the repressor of toxins (Rot).

Rot has been reported to regulate many virulence factors. This has been shown particularly in *agr*-negative mutants (Said-Salim *et al.*, 2003); since *agr* represses *rot*, these studies have shown the effects of *rot* excluding the *agr* effect. Agr has been reported to display an opposite pattern of virulence gene regulation to Rot possibly due to the

negative regulation of *rot* mRNA by RNAIII. However, the Rot-Agr relationship is more complex than this implies (Said-Salim *et al.,* 2003). It has been suggested that the regulation of Rot by Agr occurs at the post-transcriptional level (McNamara *et al.* 2000, Said-Salim *et al.* 2003, Geisinger *et al.* 2006). It is therefore thought that Rot is one of the transcriptional regulators used by Agr to control virulence gene expression. Furthermore, SarA has been reported to bind directly to the promoter region of *rot* modulating its transcription (Manna and Ray, 2007). The relationship between the Sar family members and their interaction with virulence factors is summarised in figure 1.9.

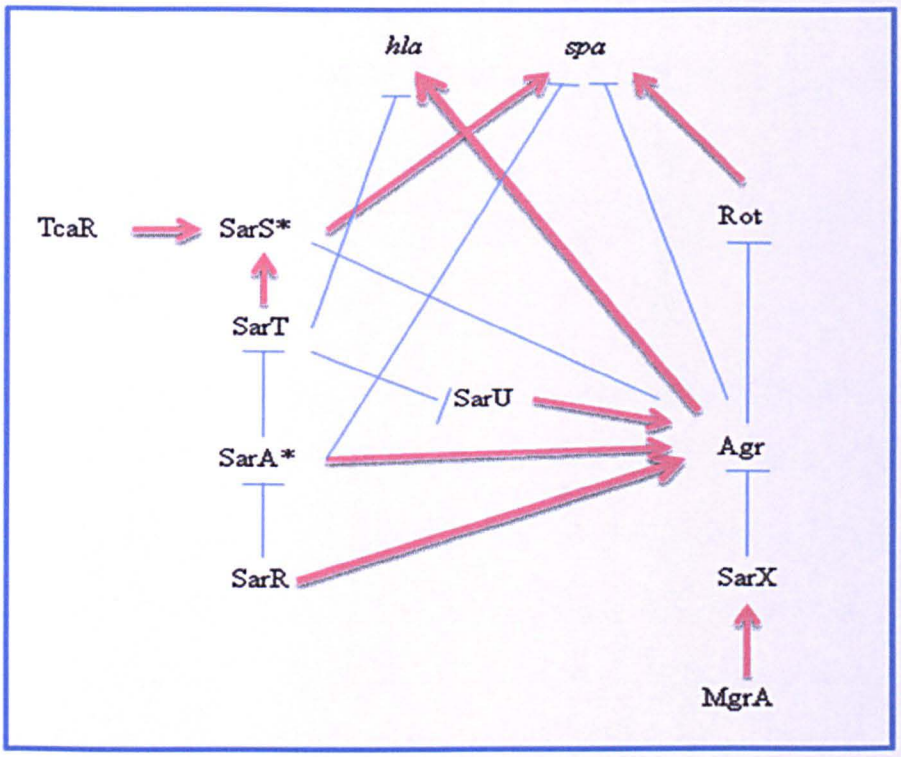


Figure 1.9 Interactions of the Sar homologues with each other and with virulence genes *spa* and *hla*. Arrows indicate positive regulation, perpendicular lines represent negative regulation, and * indicates positive regulation by σ^B. Adapted from Fournier, 2008.

1.3.2.4 The Sae system

The staphylococcal accessory element (Sae), also described as a *S. aureus* exoprotein expression locus, constitutes four genes (A-D) encoding SaeRSPQ generated from three promoters (P_{A-C}); figure 1.10. Of these four transcripts, only the functions of SaeRS are known; these two components encode the classical histidine kinase and response regulator of a two-component system.

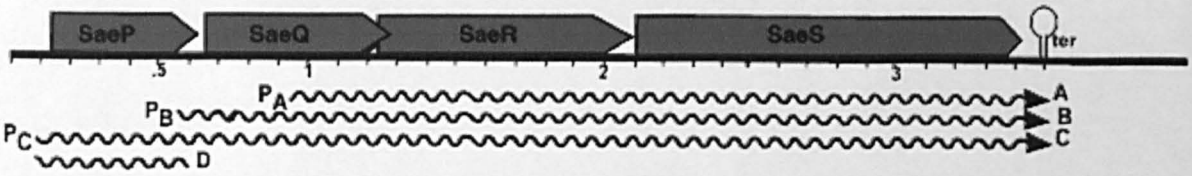


Figure 1.10 The Sae locus. Four transcripts encoding *saePQRS* are synthesised from three promoter, P_{A-C}.

The genes of this locus are also differentially expressed during growth phase; T3 is prevalent in the exponential phase, whilst the others are produced later. Agr and Sar are also required for full production of these transcripts (Novick and Jiang, 2003). The Sae system has been shown to be essential for the transcription of *hla*, *hlb* and *coa*, independently of Agr and Sar, supporting its recognition as a major regulatory system (Giraudo *et al.*, 1997). Sae upregulates *hld, seb, efb, eap, lukF, lukM, hlgACB, chp, scn, fnbA* and *fnbB* and represses *plc, sspA, aur* and *spa*.

1.3.2.5 The Arl system

Autolysins are required to facilitate cell division and separation through their role in splitting the cell wall. Furthermore, as they are involved in peptidoglycan turnover and cell suicide co-ordinated control is essential. In *S. aureus*, the autolysis related (Arl) locus,

another two-component system (Fournier and Hooper, 2000) appears to be responsible for controlling expression of these enzymes. The precise mechanisms by which this system modulates virulence gene expression have not been fully elucidated and it is supposed that other regulators may be involved (Fournier *et al.* 2001, Luong and Lee 2006). As with SarA, the Arl system has also been reported to modulate DNA supercoiling and this has therefore been proposed as a mechanism for transcription regulation (Fournier and Klier, 2004).

The Arl system transcripts are expressed during the exponential and post exponential growth phases. It is thought that the targets of ArlR could be the *sarA* regulatory factors SarR, SarA and σ^B (sigma factor B) (Fournier *et al.*, 2001). In addition Arl may affect *agr* targeted genes by reducing *agr* expression by down-regulating RNAIII transcription and thus AIP concentration. Expression of this locus is highest during the post-exponential phase. However, autoregulation is not thought to occur. SarA and Agr have been reported as positive regulators (Fournier *et al.*, 2001). In addition, MgrA (multiple global regulator, also known as NorA and Rat) and σ^B may upregulate expression but this has been reported as strain dependent (Ingavale *et al.* 2003, Luong and Lee 2006). Finally, Arl has been reported to modulate biofilm formation, although the mechanisms by which this occurs are not fully understood (Fournier and Hooper 2000, Toledo-Arana *et al.* 2005).

1.3.3 Other recognised regulatory systems

Other global regulators of virulence gene expression include σ^B (Sigma B), SrrAB (staphylococcal respiratory response, also known as SrhSR system), and LytRS.

σ^B is an alternative sigma factor that is generally activated directly within the cell rather than through signal transduction (Novick *et al.*, 2003). σ^B has been recognised for its role

in regulating stress responses and intermediary metabolism (Chan and Foster 1998, Kullik *et al.* 1998, Zeibandt *et al.* 2001, Bischoff *et al.* 2004). σ^B has been shown to be activated by environmental stress and energy depletion (reduced ATP/ ADP ratio), ethanol and salicylic acid (Chan and Foster, 1998). Regulation of σ^B occurs via a complex post-translational pathway involving *rsbU*, *rsbV* and *rsbW* (Scott *et al.*, 1999). RsbW is an anti-sigma factor that phosphorylates RsbV, an anti-anti-sigma factor. Under environmental stress, phosphorylated RsbV (Rsbv~P) is dephosphorylated by either RsbU or RsbP (phosphatases), and it is then able to bind RsbW; the result is release of σ^B and thus its activation.

The SrrAB system has been recognised as having a role in regulation influenced by environmental oxygen conditions (Throup *et al.* 2001, Yarwood *et al.* 2001). This is significant at focal points of infection (especially deep infection) where oxygen levels becomes limiting. It has therefore been proposed that SrrAB could be the link between respiratory metabolism and the expression of virulence factors (Yarwood *et al.*, 2001).

A further system, the LytRS system is involved in autolysis (Brunskill *et al.*, 1996). LytR and LytS control the expression of two transcripts LrgA and LrgB (immediately downstream of their locus) which inhibit the extracellular activity of murein hydrolases. These are enzymes which catalyse the cleavage of bacterial structural cell wall components. Thus, inhibition of these autolysins prevents cell division.

Finally, virulence factors whose primary role is not regulation have been reported to function as regulators. Included in this group are several Clp proteins such as ClpX (Frees *et al.* 2003, Frees *et al.* 2005), ClpP (Frees *et al.* 2003, Frees *et al.* 2005, Michel *et al.* 2006), ClpC (Frees *et al.* 2004, Becker *et al.* 2001), SvrA (staphylococcal virulence regulator; Garvis *et al.* 2002), Msr (methionine sulfoxide reductase; Rossi *et al.*, 2003),

CvfAB (conserved virulence factors; Matsumoto *et al.*, 2007), CcpA (catabolite control protein A; Warner and Lolkema 2003, Seidl *et al.*, 2006), Aconitase (Somerville *et al.*, 2002) and Msa (modulator of SarA; Sambanthamoorthy *et al.*, 2006). This list is expanding continually.

In conclusion, knowledge of this complex network has helped our understanding of the co-ordinated manner in which genes are temporally expressed. It is thought that it is the complexity of the pathways involved and the specificity of the interactions that facilitate survival of this organism since they enable accurate and pleotropic regulation of the expression of virulence factors in response to environmental signals.

1.4 MICROARRAY TECHNOLOGY

1.4.1 Background and principles

Microarrays are essentially a large scale dot blot, used for multiplex reactions to define the genome compositions of samples (Wildsmith and Elcock, 2001). They represent a continuation of molecular biology hybridisation methods, but in a format that allows a huge increase in the number of probes available for parallel testing. The principle of all 'arrays', as recognised in different technological platforms, is the composition of a set of pre-defined nucleic acid probes, immobilised at specific X, Y coordinates on a surface of choice. When used for typing purposes, labelled (target) nucleic acids are exposed to the probes, resulting in hybridisation in the presence of complementary sequences. This method of comparing genomes using arrays has been classified as *Genomotyping* (Lucchini *et al.*, 2001).

The key part of all microarray processes is the hybridisation step. This property of exploiting the preferential binding of complementary single-stranded nucleic acid sequences was first exploited experimentally during the 1960s, and became well accepted with the technique of Southern blotting (Gillespie and Spiegelman 1965, Southern 1975). However, with the development of sequencing projects, advances in miniaturisation and high density synthesis of nucleic acids on solid support, the principle of microarrays and microchip technologies emerged (Causton *et al.*, 2003). Microarrays were first developed in the pioneering laboratories of Schena and Brown (Schena *et al.*,1995). Shortly after, the first gene expression study using microarray technology was used in 1997 (DeRisi *et al.*, 1997). Microarray technology has changed the experimental landscape of expression analysis from low throughput (typically by northern blot analysis) to high throughput parallel analysis, where many genes can be simultaneously interrogated.

With the sequencing of several complete genomes, a vast amount of information has become available. However, the value of this information will only be fully realised when the function and expression of individual genes has been elucidated. Microarrays are greatly facilitating this process. Analysis of genomic content by microarrays provides an additional means to determine the functions of uncharacterised open reading frames. The ability to use partially characterised sequences and the flexibility of microarray design have advanced the utility of arrays for understanding many biological entities (Nubel *et al.* 2004, Troesch *et al.* 1999, Khodakov *et al.* 2008, You *et al.* 2008, Bøving *et al.* 2009). The identity of the probe sequences determines the information obtained from the array experiments, which in turn is completely dependent on genome sequencing information. Therefore, obtaining accurate information is crucial, as any deficiencies in this will impact on all downstream processes.

1.4.2 Technological overview

1.4.2.1 Microarray Production

Each feature on an array represents a nucleic acid sequence called a spot or probe. Probes can be produced from cDNA, PCR products, or from pre-synthesised or *in situ* synthesised oligonucleotides (single-stranded DNA segments). Several different platforms are available for the accurate deposition of probes onto solid support. Standard robotic arrayers are capable of spotting >20,000 probes on standard low-fluorescence microscope slides. This process is highly automated. Probes (either synthetic oligonucleotides or PCR amplicons) are spotted onto glass slides usually in a grid format using either a pin-based robotic arrayer or an ink jet microdispensing liquid handling system (Hardiman, 2004). The pin based arrayers deposit the DNA solution (picolitre quantities) by direct contact

between the pins and the solid support. Available contact printing robots use either split- or solid-pins. Split pins have a reservoir that can hold nanolitre quantities of spotting solution, allowing many spots to be printed without refilling. Solid pins can produce only a single spot before refilling unless used in the 'pin and ring' format. Non-contact printing is said to more accurately produce even spots and uses technology based on that found in ink jet printers. However, it has been reported that cross contamination is more of an issue with non-contact printing.

Spotted arrays can be fabricated on site using equipment that is now available within most medium to large size research establishments. The advantages of these platforms are that the arrays may be printed at a relatively low cost, and provide the flexibility to enable fabrication of arrays with many different designs. Unlike the commercially available microarrays comprised of oligonucleotides synthesised *in situ* (i.e. Affymetrix™ and NimbleGen™) which often accommodate only one sample per array, spotted arrays are generally used to compare two samples, each labelled with a different fluorescent dye and hybridised in parallel (Hardiman, 2004).

Very high density microarrays (Causton *et al.*, 2003) comprise oligonucleotides assembled *in situ* and are often referred to as 'chips', the term coined by the leader in the field, Affymetrix™. Synthesis involves photodependent deprotection, and physical masks (four per nucleotide position making 80 for an array of 20mers) are needed to direct light to specific areas of the chip (Lipshutz *et al.*, 1999).

The microarray market production has matured over recent years and several competing platforms are commercially available. The pioneering platform amongst these was the Affymetrix™ GeneChip. Probes consist of short oligos (~25-mers) built up by chemical synthesis by a combination of photolithography and solid-phase DNA synthesis (Chee *et*

al., 2004). These arrays are supplied pre-fabricated, therefore reducing the time associated with producing and optimising an in-house system. They have become well established in the field of transcription profiling due to their extensive genetic content and high reproducibility. Furthermore, the quality control measures already in place by the manufacturers provide an increased level of assurance for the user. This type of system is very useful for small scale experiments where flexibility in chip design is not necessary. Presently, it represents the most cost effective way of screening the genome of interest.

One of the major advantages of the Affymetrix™ GeneChips is probe redundancy. Several independent oligos binding to different regions of the same RNA are included in the design. This minimises cross-hybridisation effects, improving signal-to-noise ratio, as well as the dynamic range of detection. Up to 400,000 oligos can be supplied on a single array representing up to 10,000 genes, with 40 oligo features per gene. Recent advances in this platform have limited the probes per gene to 11 probe pairs, while at the same time reducing the feature size to as small as 18μm. Furthermore, mismatch control probes (which contain a single base difference in a central location) are also employed to control for non-specific hybridisation. These are used to subtract background and cross-hybridisation (Hardiman, 2004).

Agilent™, a competitor of Affymetrix™, also synthesise probes *in situ* but using ink jet printing via phosphoramidite chemistry (Agilent™ additionally produce spotted cDNA arrays from PCR amplicons). The oligos consist of 60-mers which are more than double the size of the 25-mers used by Affymetrix™. These longer oligos are used to overcome the poor hybridisation problem associated with shorter probe sequences. That is, larger probes provide a greater surface area for hybridisation, which subsequently provide better sensitivity. In addition, the longer probes are better able to tolerate mismatches, providing a better system for the analysis of highly polymorphic regions (Hardiman, 2004).

Compared with the Affymetrix™ GeneChips, the Agilent™ system only requires one 60-mer per gene or transcript. Finally, although both Agilent™ and Affymetrix™ also both offer user-defined custom arrays, the cost is still much higher than producing custom arrays using a standard arrayer.

The CodeLink™ Bioarray platform produced by Amersham Biosciences provides yet another commercial array. Short oligonucleotide probes (30-mers) are applied to a three-dimensional polyacrylamide gel matrix. The idea behind this technology is that the hydrophobic surface favours only specific binding, thereby minimising background noise. In addition, as with the Agilent™ system, only one oligo probe per gene is used. This oligo is empirically selected from similar probes on the basis of the signal intensity and specificity achieved with a panel of tissues (Hardiman, 2004).

Alternatives to *in situ* array fabrication (not based on conventional photolithography masks) are provided by NimbleGen™ and Febit™, which use digital mirror devices (DMD) and digital light processors (DLP) respectively. The DMD technology used by NimbleGen™ enables rapid production of custom high-density arrays in a cost effective manner. In contrast, the Febit™ system (DLP) uses a benchtop instrument to generate arrays within a three-dimensional microstructure. Four individual channel-like chambers within the 3D microstructure allow four array experiments to be run in parallel (Hardiman, 2004). The list of commercially available platforms is extensive and it seems likely that 3D arrays represent the next trend in microarray platforms. The main features of the above mentioned platforms have been summarised in Table 1.12.

Table 1.12 Comparison of microarray platforms. Taken from Hardiman. 2004.

	CodeLink™	Affymetrix™	Agilent™	NimbleGen™	Febit™	Spotted Arrays
Array format	30-mer	25-mer	60-mer	24-mer	25- to 30-mer	Variable
Hybridisation time	18h	16h	17h	16h	16h	17h
Hybridisation temperature	37°C	45 °C	60 °C	45 °C	45 °C	60 °C
Sensitivity	1:900,000	1:100,000	1:100,000	-	1:100,000	1:300,000
Advantages	Sensitivity; 3D surface; liquid hybridisation kinetics; can be utilised with any microarray scanner; customization is possible	Reproducibility; content; mature platform; customization	Reproducibility; content; mature platform; customization	Customization	Customization	Inexpensive; can be utilized with any microarray scanner; customization is possible
Disadvantages	Non-contact printing: printing-related issue, such as poor spot morphology	Short oligonucleotides: less sensitive	Two-colour dye bias and ozone-related degradation	Short oligonucleotides: less sensitive	Short oligonucleotides: less specific	Two-colour dye bias and ozone-related degradation; poor reproducibility; reporter feature identity errors

1.4.2.2 Experimental process

Fluorescent dyes have been adopted as the primary labelling method in microarray analysis. This approach offers the advantage of simultaneous detection of two or more signals in one experiment, enabling microarray users to subject two differentially labelled samples to the same array for comparative analysis (whether mRNA for expression studies, or DNA for genomic analysis). As such, the throughput of arrays has increased over filter-based macroarrays where only one radioactively labelled sample could be analysed at any one time. However, for some experimental designs, the hybridisation of a "single-labelled population of nucleic acids" to the array is still the preferred method. Here comparisons between samples are obtained from different arrays.

During the hybridisation reactions, duplexes are formed between the single-stranded immobilised probes and the denatured targets. The probes are immobilised in excess so as not to limit the kinetics of hybridisation. Equal amounts of labelled sample are combined and hybridised to the array. The fluorescence level of each spot is measured independently for each dye, post hybridisation. The most frequently used dyes for two-colour analyses are the cyanine dyes Cy3 and Cy5 which, when scanned in false colour, appear green and red respectively. Therefore, if the nucleic acid from sample 1 (labelled with Cy3) is present on an array, the probes fluoresce green. Similarly if the nucleic acids from sample 2 (labelled with Cy5) are present, the probes fluoresce red. However, if the nucleic acid from both samples are present on the same array, the probes fluoresce yellow, and if neither are present, the probes do not fluoresce (appear black when scanned). A detection system quantifies these fluorescence signals. Since each gene can be identified by its location on the slide, an output file from the scanner is generated for the relative signal abundance of each gene.

The complexity of the hybridisation process is greater than that encountered in any other molecular biology technique. The quality of data obtained from microarrays is highly dependent on the specificity of the system, as determined by the stringency of hybridisation and washing conditions. Stringent hybridisation conditions should ensure that only sequences with high homology bind to the probes on the array. Computerised data processing is a pre-requisite in all microarray data analyses due to the large amount of data generated. Often microarray scanners come with specialised software packages to extract primary information from the scanned image. Other packages may then be used for more downstream manipulation and analysis, such as normalisation and extraction of biologically meaningful conclusions. It is this process of extracting biologically meaningful data from the experiments that presents the greatest challenge in microarray analysis. Several factors can affect the outcome of this process including, in particular, the experimental design. A biologically sound and statistically robust experiment will facilitate extraction of useful information. For this reason, and to enable the organized compilation of microarray data, the Microarray Gene Expression Database (MGED) group have defined the minimal information about a microarray experiment (MAIME) guidelines for data submission (gene expression data). With the increased use of microarray technology, scientific research journals are becoming more concerned with conformance to MAIME guidelines. Additionally, post-hybridisation confirmatory studies (e.g. Northern blot and/or RT-PCR analysis) are also being proposed, necessitating the need for guidelines here. Yet, the main issues with confirmatory studies lie in the 'subjective' manner in which researchers select the genes to confirm. Furthermore, certain techniques (e.g. biopsy or laser-captured micro dissection material) produce insufficient RNA extracts for both array analysis and confirmatory studies. However, the issues raised here are concerns of whether confirmatory studies are necessary for such experiments (Rocket, 2003).

1.4.3 Advantages and disadvantages of microarrays

The rapid emergence of microarrays as a general microbiology analytical tool is reflective of its versatility. This technique has been used in a diverse range of biomedical disciplines to understand the biology of host pathogen interactions. Microarrays therefore provide insights into the dynamics of the genome being studied, as well as genomic shifts, an area recognised as functional genomics.

Microarrays primarily provide a screening tool to identify putative genes of interest. The main advantage of this technology is the ability to obtain information on many hundreds (or thousands) of data points simultaneously. With respect to gene expression studies, all the data are obtained under the same experimental conditions, eliminating this source of error in the analysis (Bunney *et al.*, 2003). At the genomic level, microarrays provide a method of differentiating strains by genotypic characterisation. Yet, since probe sequences represent specific phenotypic characteristics, microarrays also provide a link between genotypic and phenotypic characterisation (van Ijperen and Saunders, 2004). Furthermore, as advances in biological understanding come as much from DNA sequence data as well as gene annotation, microarrays provide a means to understand the function of non-annotated sequences by functional analysis. Since the principle of gene expression is that genes up- and down-regulated in the same way could have similar functions, by analysing gene expression under different conditions, the pathways in which these hypothetical genes associate may be better understood. That is, microarrays help fill in the information gap in sequencing data. Importantly, due to the high cost associated with more advanced techniques such as genome sequencing, arrays provide a rapid, cost-effective alternative. Although arrays are very much dependant on the information provided by sequencing projects, in practical terms, they provide a shortcut to whole genome comparisons.

Genomes are compared gene by gene rather than base by base, undercutting the cost associated with sequencing.

In contrast to the many advantages of using microarrays, there are also several limitations. Some of these limitations, together with possible solutions are as follows. Initial microarray setup (within a laboratory) requires a considerable amount of hands-on-time. Sample preparation can also be time consuming. However, commercial companies have realised this and some automated platforms are available. The most time consuming aspect of this technique remains data analysis (van Ijperen and Saunders, 2004). Additionally, microarray results can easily be confounded by the process of cross-hybridisation due to sequence similarities in either probes or genes (e.g. gene families). This can be limited somewhat during probe design by selecting gene sequences that are unique between family members (i.e. probe sequences based on regions that differ between closely related genes; Bunney *et al.*, 2003). Finally, as with all technologies, there is a limit of detection. With expression studies (often where obtaining large amounts of mRNA can be the limiting factor) low abundance genes may not be easily detected. To overcome this, commercial companies have developed RNA amplification techniques. However, many presently on the market cater for eukaryotic RNA amplification rather than that of prokaryotic origin.

1.4.4 Applications of microarray analysis: Comparative genome hybridisation (CGH) versus transcription profiling

CGH studies are used to detect DNA sequences in a genome, and thus the presence of genes similar to the strains used for probe design. This method of analysis is useful for detecting genomic shift between strains. CGH comparisons enable the identification of 'interesting genes' by comparing two or more different phenotypes. The aim: to provide a

possible correlation between gene carriage and disease transmission/presentation (van Ijperen and Saunders, 2004).

The sequencing projects of many bacterial pathogens have illustrated the commonalty in all genomes; that is, the presence of a core set of genes, together with the variable 'divergent' genes which are present only in some strains. Indeed what is shown is the variation in the degree of genomic plasticity. The divergent genes which are generally either located on or very close to chromosomal elements (e.g. transposons, bacteriophage, insertion sequences, or genomic islands) are lost or acquired as a unit. In more recent years, some of the more informative arrays have been composite arrays comprising probes from the 'divergent' genes in various key strains rather than strain specific probes (Witney *et al.,* 2005, Saunders *et al.,* 2004). Since all array comparisons are relative to strain specific probes used, composite probes provide greater insights and thus a more informative genotype (van Ijperen and Saunders, 2004).

Gene expression studies (transcriptomics) identify mRNA transcripts (sometimes represented as cDNA) in a genome. Transcription studies provide insights into the functional behaviour of the genome (genome dynamics) by determining which genes are induced or repressed in response to an environmental stimulus. The underlying hypothesis with such studies is that genes with similar expression levels (under the same conditions) are likely to have similar biological functions.

Traditional methods of analysing gene expression include Northern blotting, reverse transcriptase PCR and nuclease protection assays. However, these methods are only suited to analysing a small number of genes and samples at any one time. Essentially microarray methodology represents the reverse of Northern blotting principles. The target is derived from the mRNA and the probe sequences are immobilised onto the solid support. Between

Northern blotting and microarrays, filter-based gene expression developed. However, the filters were large in size, and often had the associated problem of auto-fluorescence, limiting the efficiency of this multiplex reaction.

During typical expression analysis, two samples are compared. One of these is the control, whilst the other is derived from cells subjected to a different stimulus. The presence of mRNA serves as an indicator for gene expression. Furthermore, the amount of mRNA is used as a measure for expression level. All expression levels are relative, and not absolute. This is because the intensity of the fluorescence signals not only reflects the number of hybridised fragments, but also the labelling density of each fragment, and the length of the fragments. Since the labels are incorporated during reverse transcription, a longer transcript could easily contain more labels than a shorter fragment. Thus, a strong hybridisation signal does not necessarily represent a highly expressed gene. Normalisation controls for this variation.

It has been noted that many genes are constitutively expressed, and their regulation takes place at the translational or post-translational level. The relationship between transcription and translation is not as simple as originally perceived by the 'central dogma' theory. Indeed the correlation between gene and protein expression is relatively poor. For this reason, global proteome analysis may be more informative on the phenotype than transcription analysis (Wildsmith and Elcock, 2001).

1.5 THESIS AIMS AND OBJECTIVES

1. *S. aureus* microarray and protocol development (Chapter 3.0)

i. To update the existing microarray to include further virulence-associated genes. The microarray will be used to investigate the natural population structure of *S. aureus* strains and better understand their potential to cause disease.

The microarray will be used for two main types of studies, (i) comparative genome hybridisation (CGH), and (ii) transcription profiling. CGH involves analysis of samples at the level of DNA, and therefore provides a profile of the presence and absence of genes on the array. Effectively, this highlights genetic differences between the strains under investigation. Transcription profiling, however, consists of analysis at the level of RNA and provides valuable information (implicitly) about the function of genes through analysis of their pattern of expression or regulation under particular conditions.

ii. To develop protocols for CGH and transcription profiling studies.

2. CGH studies (Chapter 4.0)

The underlying aim of this project is to use the virulence-associated microarray for comparative studies of a panel of internationally recognised healthcare-associated (HA-) and community-associated (CA-) strains. The objectives are:

i. To study variation amongst lineages of MRSA.

The array data will be used to explore genotypic differences that may explain variations in epidemiology and pathogenicity of recognised successful/ epidemic/ pandemic strains. The predominance of certain lineages (e.g. epidemic MRSA-15/-16) suggests a greater pathogenic potential. The search for differential markers will provide insights into traits which may be important in the success, transmission and pathogenicity of different lineages.

ii. To provide evidence for recombination events to gain insights into the evolutionary history of *S. aureus* via this mechanism.

CGH will be used to find strains that are closely related to the ancestral clones that we hypothesise recombined to produce the lineage EMRSA-15 (ST22-SCC*mec*IV), the most prevalent epidemic HA-MRSA in the UK. The data will be examined to identify other large-scale recombination events that may have taken place. The hypothesis is that large-scale recombination events may have had an impact on the way epidemic strains survive in hospital environments and colonise the human nasopharynx.

iii. To explore genotypic differences that may explain variations in phenotypes. In particular, patterns of genotypic variation that might account for differences in epidemiology and pathogenicity of the CA and HA *S. aureus*.

3. Transcriptosome profiling: gene expression studies (Chapter 5.0)

i. The *agr* locus of *S. aureus* encodes a regulatory RNA molecule (RNAIII) known to influence the transcription of many virulence-associated genes. In studies using an *agr* mutant, the virulence gene expression profiles of cells of strain SH1001 will be compared during exponential and stationary phases of growth. The *agr* mutant will be tested in parallel with the unmodified parental strain (SH1000) to determine the effect of this mutation on virulence gene regulation. The hypothesis is that the ability of the knock-out mutants to alter their transcriptional activity in response to stress or quorum sensing signals will be impaired.

ii. Compare expression profiles of cells grown under biofilm simulating conditions to their planktonic counterparts in order to determine putative biofilm-associated virulence genes.

2.0 MATERIALS AND METHODS

2.1 Description of *S. aureus* strains

Forty three strains (26 healthcare-associated and 17 community-associated strains) were provided by the Laboratory of Healthcare-Associated Infections (LHCAI), Centre for Infections, Colindale. All strains were typed by MLST and PFGE, and SCC*mec* type was defined by PCR. Additional information (*agr* type, *spa* type, and toxin profile) was provided for some of the strains (Table 2.1). Isolates were classified as healthcare-associated or community associated from molecular and epidemiological data. In addition, a collection of 17 epidemic MRSA (EMRSA) isolates were received from the National Collection of Type Cultures (NCTC) via LHCAI. Strains were provided on nutrient agar by LHCAI, and later stored on beads (Pro-lab diagnostics). In addition, two isogenic strains, one carrying a mutation in the *agr* (accessory gene regulator) locus and its parental progenitor, were provided by Professor Simon Foster's laboratory, University of Sheffield. Finally, one *S. aureus* strain isolated from a chronic wound at Heath Hospital (Cardiff) was used in a collaborative study with the Department of Oral Surgery, Medicine and Dentistry (School of Dentistry, Cardiff University).

Table 2.1 Strains used in this project. All strains were provided by LHCAI, unless otherwise stated (see text). Replicates indicated by '*'. Isolates were classified as healthcare-associated or community associated from molecular and epidemiological data.

Strain Identifier	MLST/ SCC*mec* Type	MLST CC	PVL	Toxin gene profile	*spa*	*agr*	H/C
Berlin	ST45-IV	45		enterotoxins G & I			H
Iberian	ST247-I	8		enterotoxin B			H
NY/Japan	ST5-II	5		enterotoxins G & I			H
Paediatric (USA800)	ST5-IV	5		enterotoxins G & I			H
South German	ST228-I	5		enterotoxins G & I			H
Veterinary Strain	ST398	398		none detected			H
Irish-1 (var)	ST8-II	8		enterotoxin A			H
Irish-2 (var)	ST8-IV	8		none detected			H
EMRSA 15 var B3 (0412)	ST22-IV	22		enterotoxins G & I		1	H
EMRSA 15 var B3 (0414)	ST22-IV	22		enterotoxins G & I		1	H
EMRSA 15 var B3 (0415)	ST22-IV	22		enterotoxins G & I		1	H
USA400	ST1-IVa	1	+	enterotoxins A & H, PVL	t127	3	C
ST5-IV PVL	ST5-IVc	5	+	enterotoxins D,G,I &J, PVL	t002	2	C
USA300	ST8-IVa	8	+	PVL	t008	1	C
S.W. Pacific clone (USA1100)	ST30-IVc	30	+	enterotoxins G & I, PVL	t019	3	C
S.E. Asia clone (USA1000)	ST59-V	59	+	enterotoxin B, PVL	t437	1	C
European Clone	ST80-IVc	80	+	enterotoxin H, exfol D & PVL	t044	3	C
MRSA252	ST36-II	30					H
MSSA476	ST1	1					C
NCTC 8325	ST8	8					C
COL	ST250-I	8					H
Mu50	ST5-II	5					H
N315	ST5-II	5					H

Strain Identifier	MLST/ SCC*mec* Type	MLST CC	PVL	Toxin gene profile	*spa*	*agr*	H/C
WA-MRSA	ST1-Iva	1		enterotoxins A & H	t127	3	C
ST772-V/ Bengal-Bay clone	ST851-V	1	+	enterotoxins A, G & I, PVL			C
Queensland clone	ST93-?IV	93	+	PVL			C
ST866-IV	ST866-IV	5	+	enterotoxins A, C, G & I, PVL, TSST			C
ST88-IV	ST88-IVa	88	+	PVL	t690	3	C
ST97	ST97-V	97		enterotoxins D & J			C
ST5-IV neg	ST5-IV	5		enterotoxins C,D, G, I & J, TSST-1			C
ST22-IV	ST22-IVc	22	+	enterotoxins G & I, PVL	t005	1	C
EMRSA-15 variant B1	ST22-IV	22		enterotoxins C, G & I			H
EMRSA-15 variant B3	ST22-IV	22		enterotoxins G & I			H
EMRSA-15 variant B5	ST22-IV	22					H
EMRSA-15 variant B7	ST22-IV	22					H
EMRSA-15 variant B27	ST22-IV	22					H
EMRSA-16 variant A1	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-16 variant A2	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-16 variant A14	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-16 variant A16	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-16 variant 29 (r1) *	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-16 variant 29 (r2) *	ST36-II	30		enterotoxins A, G & I, TSST-1			H
EMRSA-1	ST239-III	8		enterotoxin A	t037	1	H
EMRSA-2	ST8-IV	8		enterotoxin A (variable)	t190	1	H
EMRSA-3	ST5-I	5		enterotoxins G & I	t001	2	H
EMRSA-4	ST239-III	8		enterotoxin A	t037	1	H
EMRSA-5/Iberian clone	ST247-I	8		enterotoxins A & B	t051	1	H
EMRSA-6	ST8-IV	8		enterotoxin A (variable)	t190	1	H
EMRSA-7	ST239-III	8		enterotoxin A	t037	1	H
EMRSA-8	ST250-I	8		none detected	t194	1	H

Strain Identifier	MLST/ SCC*mec* Type	MLST CC	PVL	Toxin gene profile	*spa*	*agr*	H/C
EMRSA-9	ST240-III	8		none detected	t037	1	H
EMRSA-10/Hannover clone	ST254-IV	8		enterotoxins A & B	t009	1	H
EMRSA-11	ST239-III	8		enterotoxin A	t037	1	H
EMRSA-12	ST8-IV	8		enterotoxin A (variable)	t190	1	H
EMRSA-13	ST8-IV	8		enterotoxin A (variable)	t190	1	H
EMRSA-14	ST8-IV	8		enterotoxin A (variable)	t190	1	H
EMRSA-15	ST22-IV	22		enterotoxins C, G & I	t022	1	H
EMRSA-16 (Tk1) *	ST36-II	30		enterotoxins A, G, I & tsst	t018	3	H
EMRSA-16 (Tk2) *	ST36-II	30		enterotoxins C, G & I	t018	3	H
EMRSA-17	ST247-I	8		enterotoxin A	t051	1	H
ST779-IV	ST779-IV			exfoliative D			H
ST39-II	ST39-II	30		enterotoxins C, G and TSST-1			H
MW2	ST1-IV	1	+				C
D76							
Agr $^{+}$							
Agr $^{-}$							

2.2 Transcription profiling

2.2.1 Culture of *S. aureus* cells for transcription profiling analysis

2.2.1.1 Standard liquid culture

Isolates were recovered from frozen beads by sub-culturing onto nutrient agar overnight at 37°C. A single colony was inoculated directly into 10ml LB broth (HPA) and incubated (37°C) overnight with shaking. An aliquot of the culture was used to make a fresh suspension with an OD_{600} reading of 0.005 (50ml final volume) and grown with shaking at 120rpm (37°C). The purity of the inoculum was tested by plating out a loopful of the starter culture. Cells were removed at different time points determined by growth curve analysis (section 3.1.2.1.2). Samples of the culture were stabilised in two volumes of RNAprotect bacteria reagent (Qiagen), and collected by centrifugation (15min, 3,000 x g). Cell pellets were either used for RNA extraction immediately or stored (-20°C) until later use.

2.2.1.2 Biofilm study: constant depth film fermenter

Biofilms were grown aerobically at 37°C in a constant depth film fermenter (CDFF) by the Department of Oral Surgery, Medicine and Dentistry. The CDFF instrument provides a simplistic, reproducible model to grow biofilms of fixed thickness. The instrument consists of a glass vessel housing a turntable containing 15 polytetrafluoroethylene (PTFE) pans. Each pan contains five recessed PTFE plugs (5mm diameter) on which the biofilms form. After inoculation, sterile media flows over the biofilms, which are maintained at a constant depth of 15 microns. Excess growth is removed by rotating PTFE scraper blades.

Cells from five to seven pans were stabilized in two volumes RNAprotect (Qiagen), and transported to the Centre for Infections (HPA) on dry ice.

2.2.2 Growth curves of *S. aureus* strains

Growth curves were used to estimate growth phases and determine the best time points for RNA extraction. One millilitre samples were removed from liquid broth cultures at prescribed times (0h, 1h, 2h, 3.5h, 4.5h, 5.5h, 6.5h, 7.5h, 16h) for quantification. In a sterile microtitre plate, serial 10 fold dilutions (down to 10^{-7}) were made in PBS containing 0.1% triton X100 (Severn Biotechnology, UK). Twenty microlitre aliquots of each dilution were transferred onto nutrient agar plates (HPA) in triplicate as described by Miles and Misra (1938). Plates were incubated overnight at 37°C and counted on the following day. Growth curves were drawn in Microsoft Excel.

2.2.3 Quantification of cells in culture

To estimate cell numbers rapidly, one millilitre cell suspensions were analysed using the Densimat (Biomerieux), a photometric reader that counts cells on the McFarland standard scale.

2.2.4 RNA isolation

2.2.4.1 RNeasy mini (Qiagen)

Cell pellets were resuspended in 100µl of lysozyme/lysostaphin solution containing 2X TE (Sigma, UK), 0.3mg/ml lysozyme (Sigma UK), 0.03mg/ml recombinant lysostaphin (Sigma UK) and 1.2% Triton X100 (Severn Biotechnology, UK) then incubated at 37°C

for 30-40min. RNA was then extracted using the RNeasy mini kit (Qiagen) from the lysis step onward.

Briefly, cells were lysed in 350µl RLT Buffer containing β-mercaptoethanol, with thorough mixing by vortexing. Two hundred and fifty microlitres of ethanol (Sigma) was added to the lysate and mixed by gentle pipetting, before applying to an RNeasy mini column. RNA was bound to the column by a brief spin (15sec, 10,000 x g), followed by a wash step with 350µl Buffer RW1. DNA was removed by the recommended on-column digestion step; 10µl DNaseI and 80µl RDD Buffer mix was incubated on the column for 15min at room temperature. The enzyme was removed by a further wash step with 350µl Buffer RW1 with brief centrifugation (15sec, 10,000 rpm). The column was washed twice more in 500µl each of Buffer RPE, and the column dried by centrifuging for 2min at full speed. Finally, the RNA was eluted in a sterile tube with 30-50µl RNase-free water. RNA (from all protocols) was stored at -20°C.

2.2.4.2 RNeasy Midi (Qiagen)

The RNeasy Midi protocol was identical to the mini protocol except that larger volumes were used enabling more cells to be extracted from a larger culture. The RLT buffer and ethanol volumes were increased by 5.7 and 5.6 times respectively, and the Buffer RPE by a factor of 5. The volume of DNase solution was doubled, and finally the elution volume was increased to 150µl.

2.2.4.3 PureYield RNA Midi (Promega)

Cell pellets that had been collected from RNAprotect solution were resuspended in 100µl lysozyme/lysostaphin solution for 30min at 37°C. Cells were lysed in 1ml Lysis Solution

containing β-mercaptoethanol, mixed thoroughly and incubated on ice for 10min. Two millilitres of the lysate were transferred to a sterile 1.5ml tube, diluted in RNA Dilution Buffer, and mixed immediately by inverting the tube several times. One millilitre of Clearing Agent was added to the tube, and again mixed by inversion. The tube was then incubated at 70°C for 5min, cooled to room temperature, homogenized with shaking and applied to the Clearing Column. Sample impurities were bound to the column by centrifuging for 10min (2,000 x g), and the cleared lysate (flow-through) transferred to a sterile 50ml falcon tube. Isopropanol (4ml) was added to the cleared lysate, mixed by inversion, applied to the Binding Column and centrifuged for 10min. The flow-through from this column was discarded and the column was washed twice in 20ml and 10ml volumes of RNA Wash Solution (5min, 2,000 x g). The column was dried by centrifugation (5min, 2,000 x g). Finally, the RNA was eluted in 1ml nuclease-free water.

2.2.4.4 PureLink Micro to Midi (Invitrogen)

Cell pellets were resuspended in 100µl lysozyme/lysostaphin solution (see Section 2.2.4.1) and mixed with 0.5µl of 10% SDS (Gibco). The mixture was incubated for 5min at room temperature then 350µl RNA Lysis Solution (containing β-mercaptoethanol) was added. The tube was mixed thoroughly by vortexing to ensure complete lysis. Two hundred and fifty microlitres of 100% ethanol (Sigma) was added and mixed thoroughly. The sample was then transferred to the RNA Spin Cartridge and centrifuged briefly (15sec, 12,000 x g) at room temperature. At this point DNase treatment was performed. This involved an initial wash step with 350µl Wash Buffer I, brief centrifugation (5min, 12,000 x g) and incubation (15min at room temperature) with 10U DNaseI in DNase buffer (80µl volume). Finally, the cartridge was washed with a further 350µl Wash Buffer I to remove the DNase. Following DNase treatment, the cartridge was washed twice in 500µl of Wash Buffer II

containing ethanol, centrifuged briefly (15sec, 12,000 x g), and dried with a final spin. RNA was eluted in 30-50µl RNase-free water.

2.2.4.5 RNAspin Mini (GE Healthcare)

Cell pellets containing 5 x 10^9 cells (prepared from a quantified liquid culture) were lysed by the addition of 350µl Buffer RA1 and 3.5µl β-mercaptoethanol, with thorough mixing. Each lysate was applied to an RNAspin Mini Filter and centrifuged for 1min at 11,000 x g. The filtrate was transferred to a sterile 1.5ml tube to which 70% (v/v) ethanol (350µl) was added. The preparation was vortexed and applied to an RNAspin Mini column. The mixture was pipetted 2-3 times and centrifuged briefly (30sec, 8,000 x g) to bind RNA. Three hundred and fifty microlitres Membrane Desalting Buffer was then applied to the membrane and spun for 1min. According to the manufacturer, this step improves the efficiency of DNase digestion.

A DNase reaction mixture was prepared by adding 10µl reconstituted DNaseI to 90µl DNase reaction buffer. Of this mixture, 95µl was applied directly to the silica membrane and incubated at room temperature for 15min. The column was then washed three times in 200µl, 600µl and 250µl each of Buffer RA2 and twice Buffer RA3 respectively, with 1min centrifugations (11,000 x g). The column was dried with a further 2min centrifugation, and finally eluted in two successive volumes of 50µl RNase-free water.

2.2.5 RNA enrichment

Ribominus® Bacteria Transcriptosome Isolation Kit (Invitrogen) for the removal of rRNA

The Ribominus® kit is based on the principle of removing ribosomal RNA (rRNA) from total RNA using specific probes which bind to magnetic beads. During the initial step, the beads are prepared for hybridization. This is followed by the hybridization of the rRNA specific probes with their target in the sample. Finally, the rRNA probe complex is separated from the sample using the magnetic beads. The concentration module is then used to concentrate the remaining sample into a small volume. The details of this method are as follows.

The Ribominus® Magnetic beads were prepared by mixing the beads thoroughly in their bottle, prior to aliquoting a 250µl suspension into a sterile 1.5ml tube. The beads were separated from their solution on a magnetic stand and the supernatant discarded. The beads were then washed twice in 250µl RNase-free water, each time discarding the suspension solution by pelleting the beads using the magnetic stand for aided separation. This was repeated for 250µl Hybridization Buffer B10, before finally re-suspending the beads in 100µl Hybridization Buffer B10. The suspension was kept at room temperature during sample preparation.

For hybridization, total RNA (2-10µg) was combined with 400pmol Ribominus™ probe, and 100µl Hybridization Buffer B10, up to a final volume of 124µl. The mixture was incubated for 5min at 37°C and snap cooled on ice. Next, the 124µl sample mix was combined with the bead suspension, mixed by thoroughly vortexing and incubated at 37°C for 15min, with occasional gentle mixing. The tube was then placed on a magnetic stand

to pellet the rRNA-probe complex. The supernatant containing the Ribominus™ RNA fraction was then transferred to a sterile tube for concentration using the Ribominus™ concentration module.

For RNA concentration, 250µl Binding Buffer (L3) and 125µl ethanol (100%) were mixed thoroughly with the supernatant, and applied to a column. The column was centrifuged (12, 000 x g) for 1min and the flow through discarded. The column was washed twice in 200µl each of Wash Buffer B5 and centrifuged (12, 000 x g) for 1min each and a further 2-3min to dry the column. Finally, the column was placed in a sterile 2ml collection tube and the RNA eluted in 10-15µl RNase-free water. RNA was stored at -20°C.

2.2.6 RNA quality check

2.2.6.1 Standard agarose gel electrophoresis

The integrity of the RNA and/or DNA samples was checked by horizontal gel electrophoresis on a 1% (w/v) agarose gel in TBE buffer. Agarose gels (Invitrogen) were prepared in Ultrapure 1 X TBE buffer (Invitrogen), heated in a microwave for 30sec (or until dissolved) and cast into a mould. One microlitre of sample mixed with 0.5µl 5X Loading Buffer (Bioline, UK) or 5µl of the DNA Hyperladder (Bioline, UK) were loaded into wells. Samples were electrophoresed at 120V for 20-30min, or until the dye had migrated three quarters of the way down the gel. Gels were then stained in ethidium bromide (1µg/ml) in distilled water. Finally, the gel was visualised under UV light (570-640nm) using a CCD camera on an AutoChemi Imaging System (UVP).

2.2.6.2 Denaturing agarose gel electrophoresis

Briefly, RNA was dissolved in loading buffer consisting of 10µl formamide, 2µl formaldehyde, 2µl 10X TBE, 4µl RNase-free water, and 2µl bromophenol blue. Samples were denatured at 65°C for 2min and cooled on ice. Products were separated on a 1% (w/v) denaturing agarose gel in 0.5X TBE buffer. After staining with ethidium bromide the separated bands were examined under UV light as described (section 2.2.6.1).

An alternative to this method was the use of a urea buffer for denaturing RNA instead of the formamide-formaldehyde buffer. The urea buffer consisted of 7M Urea, 13% Ficoll, 0.01% bromophenol blue, and 10X TBE, up to a final volume of 18µl.

2.2.6.3 Qubit quantification system (Invitrogen)

For DNA/RNA quantification, samples were diluted 200-fold in a Quant-iT master mix solution comprised of 199 volumes of Quant-iT (dsDNA or RNA) buffer and 1 volume of Quant-iT (dsDNA or RNA) reagent. Pre-prepared standards (0ng and 10ng) were diluted 10-fold in the master mix solution, and were used to calibrate the instrument. Calibration values were measured using the Qubit Fluorimeter (Invitrogen) by measuring the fluorescence of the standards. Using these values, the fluorimeter quantifies the concentration of dsDNA or RNA in the sample.

2.2.7 *In vitro* transcription (IVT), amplification and labelling

For transcription profiling, RNA was converted to cDNA by IVT, and labelled with fluorescent dyes. A number of protocols were tested, involving IVT alone (with a later

labelling step) or IVT with labelling all in one step. Amplification of cDNA was also included in some of the protocols.

2.2.7.1 IVT: First strand cDNA synthesis

2.2.7.1.1 Random Primer method with MMuLV enzyme

Reverse transcriptase reactions consisted of total RNA (5μg), 400U MMuLV reverse transcriptase (New England Biolabs, NEB, UK), 10X MMuLV buffer (NEB, UK), 3μg Random Primer 9 (NEB, UK), 40U RNase inhibitor (NEB), dNTP mix (20mM each of dATP, dCTP, dGTP and dTTP; Invitrogen) and nuclease-free water in a final volume of 30μl. Reaction mixes were incubated at 42°C for 2.5h and cooled to 4°C for 2min. The RNA (in RNA:cDNA heteroduplex structures) was then digested by the addition of 5U RNase H in 10X RNase H buffer (NEB, UK). Complementary DNA (cDNA) was purified from the reaction mixtures using the MinElute Cleanup Kit (Qiagen, UK) with an optimised protocol.

Briefly, 300μl of buffer ERC and 30μl 3M sodium acetate (pH 5) were applied to the columns. The samples were added to the mixture on the column, mixed and centrifuged for 1min at 10,000rpm. The columns were washed in 750μl buffer PE and spun to dry. Finally, the cDNA was eluted twice in 10μl 10mM sodium-phosphate (pH 8.5). The eluates were pooled, dried in a heating block at 65°C and then resuspended in 5μl 0.1M carbonate coupling buffer pH9 (2:1 $NaCO_3$:$NaHCO_3$, v/v) and coupled with either Cy3 or Cy5 monoreactive dye (NHS ester, Amersham) in DMSO (2μl DMSO per sample with one dye pack split for up to 3 samples). Samples were mixed thoroughly, wrapped in foil and incubated at room temperature for up to 90min in the dark. Uncoupled dyes were removed

by MinElute purification (according to the manufacturers' instructions) and cDNA was recovered in elution buffer (10µl).

2.2.7.1.2 CMV-tailed nonamer method with MMuLV enzyme

This method involved a modification of the random primer method in which 3µg CMV-tailed nonamer-F (Table 2.2) was used instead of the 3µg random primer. All other reagents and conditions remained the same. The strategy was to produce first- and second-strand cDNA using the tailed-nonamers, followed by amplification using CMV-primers.

2.2.7.2 Second strand cDNA synthesis and amplification

2.2.7.2.1 CMV-tailed nonamer method (Taq polymerase enzyme) and amplification

The second strand was produced using CMV tailed-nonamer-R (Table 2.2) in a reaction volume of 30µl, consisting of 3µl 10X Taq polymerase buffer (Invitrogen), 1µl of 6mM dNTP mix, 2µl $MgCl_2$ (50mM), 3µl CMV tailed-nonamer-R (3µg), and 0.4µl (2U) Taq polymerase (Invitrogen). The reaction mixture was incubated for 3 min at ambient temperature, followed by 30 min incubation at 42°C, and 5 min at 72°C. The cDNA was then amplified using the non-tailed CMV primers (Table 2.2). Briefly, 2µl of 2mM dNTP mix was added to the cDNA, as along with 5µl each of CMV-F and CMV-R primers (5pmol/µl each). The cycle parameters were 93°C 10sec, 65°C 10sec, and 74°C 30sec. Amplified products were purified using the MinElute reaction cleanup kit (Qiagen).

2.2.7.2.2 CMV-tailed nonamer method with exo- Klenow enzyme

This method was a slight adaptation of the Taq polymerase method (section 2.2.7.2.1). Five units exo- Klenow enzyme (NEB) was used, along with its 10X buffer. Additionally, the dNTPs were at a final concentration of 2mM.

2.2.7.2.3 CMV-tailed nonamer method with Superscript enzyme

This method follows the exo- Klenow method (section 2.2.7.2.2), except that the first strand was transcribed using Superscript enzyme (400U) in 5X Superscript buffer.

2.2.7.3 IVT (Random Primer 9) and direct labelling

A mixture of 2μg total RNA and 3μg Random Primer 9 (NEB) in a final volume of 21μl was heated to 70°C for 5min and cooled to 20°C. An IVT labelling mixture (9μl) consisting of 10X MMuLV buffer (3μl; NEB), 400U MMuLV (2μl; NEB), 40U RNase inhibitor (1μl; NEB), 2μl dNTP mix (2mM each of dATP, dCTP, dGTP, and 1mM dTTP; Invitrogen) and 1μl Cy3-dUTP or Cy5-dUTP (GE Healthcare, UK)) was added to the RNA/random primer mix,. The reaction was incubated at 42°C for 2.5h to generate fluorescently labelled DNA. The RNA template was removed by adding 5U RNase H and 10X RNase H buffer (final volume 20μl) and incubating at 37°C for 20min. The labelled samples were again purified using the MinElute Reaction Cleanup kit (Qiagen) and eluted in10μl.

2.2.7.4 IVT and indirect labelling

2.2.7.4.1 Random Primer method

The initial part of the protocol for amino-allyl indirect labelling was the same as the direct labelling of cDNA (section 2.2.7.3), except that the dNTP mix included 5-aminohexylacrylamide-dUTP (aha-dUTP, Invitrogen). A 20X working stock of aha-dUTP mix was prepared containing 22.22mM each of dATP, dCTP, dGTP and aha-dUTP, and 11.11mM dTTP. Aha-dUTP is a dTTP derivative which is incorporated into cDNA and subsequently labelled with the NHS-ester derivative of a fluorescent dye (indirect method). Unincorporated nucleotides were removed using the MinElute Reaction Cleanup kit.

2.2.7.4.2 CMV-tailed nonamer method with indirect labelling

In this variation of the Random Primer method using MMuLV (section 2.2.7.1.1), the 20mM dNTP mix was replaced with 2µl of 20X aha-dUTP mix, 5X MMuLV buffer was used instead of the original 10X buffer and 3µl 0.1M DTT was added to the mixture giving a final volume of 29µl.

2.2.7.4.3 Superscript method

This method was similar to the CMV-tailed nonamer method (section 2.2.7.4.2) except that the enzyme was SuperScript III (Invitrogen) was used along with its corresponding 5X SuperScript buffer. Additionally, the concentration of the aha-dNTP mixture was increased to 2X.

2.2.7.5 Promega Chipshot kit

RNA was reverse transcribed and labelled using the ChipShot Indirect Labelling kit (Promega) according to the manufacturer's protocol (with the exception of the RNA starting concentration). Briefly, total RNA (2.5μg instead of 5μg) was reverse transcribed to cDNA by random priming in the presence of amino-allyl dNTP. The initial RNA template was then degraded, and the remaining cDNA was purified. The synthesized cDNA was then coupled with Cy3 or Cy5 mono-functional dye (NHS-ester). Finally, the labelled samples were purified to remove excess dye.

2.2.8 Real-time PCR

Real-time polymerase chain reaction (RT-PCR) was carried out in a total volume of 10μl containing 1μl sample (genomic DNA/ cDNA), 0.5pmol of each primer (Table 2.2) and 2μl LightCycler FastStart DNA MasterPLUS SYBR Green I (Roche Diagnostics). The reaction was performed as follows: HotStart Taq polymerase was activated at 95°C for 10min, followed by 50 cycles of denaturation at 92°C for 0sec, annealing at 55°C for 0sec and 58°C for 0sec, and extension at 74°C for 10sec. The ramp rate between annealing temperatures was 3°C/sec and all other ramp rates were at maximum. For quantification, the fluorescence signal associated with the binding of SYBR Green I to the double-stranded PCR product was recorded at the end of each cycle. After the PCR reaction, the temperature was decreased to 50°C and then gradually raised to 92°C at a ramp rate of 0.1°C/sec. Fluorescence was continuously monitored during this process for melting curve analysis. A negative control (water as sample) was included in each run. Primer pairs were synthesised by MWG-biotech AG (UK).

Table 2.2 Sequences of tailed random nonamers and primers (5'-3') used in the methods.

Oligonucleotide	Sequence	Use
16S rRNA forward primer	GGA TCG TAA AAC TCT GTT ATT	RT-PCR; to detect chromosomal DNA
16S rRNA reverse primer	TTC ACA TCA GAC TTA AAA AAC	RT-PCR; to detect chromosomal DNA
CMV forward (CMV-F) primer	GAG GAC AAC GAA ATC CTG TTG GGC	RT-PCR; to produce & quantify cDNA
CMV (CMV-R) reverse primer	GTC GAC GGT GGA GAT ACT GCT GAGG	RT-PCR; to produce & quantify cDNA
CMV-F tailed random nonamers	[(AGCT)x9] GAG GAC AAC GAA ATC CTG TTG GGC	cDNA synthesis
CMV-R tailed random nonamer	[(AGCT)x9] GTC GAC GGT GGA GAT ACT GCT GAGG	cDNA synthesis

2.3 Comparative genome hybridization

2.3.1 Culture of S. *aureus* cells

Bacterial cells were plated onto nutrient agar (HPA, Colindale UK) and incubated overnight for 16-18 h at 37°C. The following day a single colony was spread onto a fresh nutrient agar plate to form a dense layer for DNA extraction after overnight culture. All isolates were stored on Microbank beads (Pro-lab diagnostics) at -20°C.

2.3.2 Genomic DNA extraction

2.3.2.1 DNeasy kit (Qiagen)

Cells were harvested from nutrient agar plates with a 1µl culture loop and resuspended in 100µl lysozyme/lysostaphin solution (Section 2.2.4.1) and incubated for 30-40min at 37°C. DNA was then extracted according to the manufacturers' protocol.

Briefly, cells were lysed by the addition of 25µl Proteinase K (Qiagen, UK) and 200µl buffer AL, with vortexing to ensure complete mixing. The mixture was incubated at 70°C for 30min to complete lysis. DNA was precipitated by the addition of 100% ethanol (200µl) and applied to the DNeasy column for binding. The column was centrifuged for 1min (8,000 x g), and washed twice in 500µl each of buffer AW1 and AW2. The column was then dried by a further centrifugation (3min, full speed) before finally eluting the DNA in 200µl buffer AE.

2.3.2.2. Pitcher method (adapted)

This method was an adaptation of a DNA extraction method described by Pitcher and Saunders (1989).

An inoculation loop of bacterial cells taken from a nutrient agar culture plate were resuspended in a 125µl mixture of lysozyme/ lysostaphin as before, and incubated at 37°C for 30min. Cells were then lysed by gentle pipetting in 2 volumes of L2 buffer containing guanidine isothiocyanate and EDTA (Severn Biotechnology, Kidderminster, UK). Half the volume of 7.5M ammonium acetate was added, with mixing. A further 450µl chloroform:isoamyl alcohol (24:1) was added and the mixture was vortexed thoroughly. The phases were separated by centrifugation (10,000 x g) for 2min, and the supernatant transferred to a fresh tube. DNA was precipitated by the addition of 0.54 volumes of ice cold propan-2-ol, with mixing by inversion of the tube several times. The fibrous DNA was then removed using a pipette tip into a sterile 1.5ml tube and washed twice in 1ml 80% ethanol and air dried overnight. Finally, the DNA was resuspended in 50µl sterile deionised water (Sigma).

2.3.3 DNA amplification

2.3.3.1 GenomiPhi DNA Amplification Kit (GE Healthcare)

DNA (5ng) in sample buffer, up to a final volume of 10µl, was denatured at 95°C for 3min and snap cooled on ice. Single-stranded DNA was then combined with 10µl of the amplification mixture (containing 1µl enzyme and 9µl reaction buffer), and incubated at 30°C for 16-18h. The enzyme was heat inactivated for 10min at 65°C, and the sample cooled at room temperature before purification.

To purify the amplified DNA, 1 volume water and 0.1 volume sodium acetate/EDTA buffer was mixed with the sample. This was followed by ethanol precipitation with 100µl 100% ethanol and centrifugation at 12,000 x g for 15min. The supernatant was removed by aspiration, and the pellet washed in a large volume of 70% ethanol. DNA pellets were centrifuged (12,000 x g) again for 1min to ensure that they were adhered to the tube wall before the supernatant was removed by aspiration. Pellets were air-dried and then resuspended in TE.

2.3.3.2 Whole genome amplification with Phi29

5ng DNA in a volume of 10µl was denatured at 98°C for 30sec and snap cooled on ice. A 40µl amplification reaction was added, consisting of 5nmoles hexanucleotide primers (NEB), 10X Phi buffer (NEB), 10µg BSA (NEB), dNTP mix (10mM each; Invitrogen), 0.005U inorganic pyrophosphatase (Roche Diagnostics) and 1 unit Phi 29 DNA polymerase (NEB). The sample was mixed and then incubated at 30°C for 16-18h. The amplified products were then purified as described for the GenomiPhi DNA Amplification method (section 2.3.3.1).

2.3.4 DNA labelling methods

2.3.4.1 Direct labelling

2.3.4.1.1 Bioprime® Array CGH Genomic Labeling System (Invitrogen)

A mixture of 2µg genomic DNA and 10µl of 2.5X random primer solution, to a final volume of 21µl, was heated to 95°C for 5min and snap cooled on ice for 5min. A labelling

mixture of 3.5µl was then added, consisting of 2.5µl 10X dUTP Nucleotide Mix, 0.5µl Cy3- or Cy5-dUTP and 0.5µl exo-Klenow fragment. Tubes were wrapped in foil and incubated for 2h at 37°C. The reaction was stopped by the addition of 2.5µl Stop buffer, and the samples purified using the MinElute Reaction Cleanup kit (Qiagen).

2.3.4.2 Indirect labelling

2.3.4.2.1 Bioprime® Plus Array CGH Indirect Genomic Labeling System (Invitrogen)

A mixture of 4µg genomic DNA and 20µl of 2.5X random primer solution, to a final volume of 44µl, was heated to 95°C for 5min and snap cooled on ice for 5min. An indirect labelling mixture of 5µl of 10X nucleotide mix with amino-allyl dUTP and 1µl exo-Klenow fragment was added. Reaction tubes were wrapped in foil and incubated for 2h at 37°Cin the dark. The reaction was stopped by the addition of 2.5µl stop buffer, and the samples purified using the MinElute reaction cleanup kit (Qiagen). The purified DNA was precipitated with 10µl 3M sodium acetate, 2µl glycogen and 300µl ice cold 70% ethanol and incubated for 30min at -20°C. The DNA was centrifuged (4°C) at maximum speed for 20min, and the pellet washed in 250µl of ice cold 70% ethanol. Finally, the pellet was air dried.

For the coupling reaction, 5µl of 2X Coupling buffer and 3µl water first resuspended with the dried pellet. To this was added 2µl Alexa Fluor dye resuspended in DMSO. The tubes were mixed thoroughly, wrapped in foil, and incubated at room temperature for 1h in the dark. Finally, the labelled DNA was purified using the MinElute Reaction Cleanup kit (Qiagen).

2.3.4.2.2 Adaptation of the DeRisi cDNA labelling protocol to label genomic DNA

Genomic DNA (1μg) and random primer mixture (6μg final of 8, 9, 10mers; Invitrogen), in a final volume of 16μl, were heated to 95°C for 30sec and then snap cooled on ice. Labelling mixture consisting of 10X exo-Klenow buffer (NEB), dNTP mix (20mM dATP, dCTP, dGTP each, and 10mM dTTP and aha-dUTP; Invitrogen) and 5U exo-Klenow fragment enzyme (4μl total; NEB) were added and incubated at ambient temperature for 30min, followed by a further incubation at 37°C for 1h. Nucleic acid was purified using the optimised MinElute protocol (section 2.2.7.1.1) and labelled as described above (2.3.4.2.1).

2.4 Microarray development

2.4.1 Array design and printing

2.4.1.1 Oligonucleotide probe design

The selection of oligonucleotide probes for the array was described by Saunders *et al.* (2004). Briefly, the design of the array was intended to include *S. aureus* genes that have a potential virulence association with the intent of investigating the virulence repertoire of clinical strains. A genome browser containing the complete annotations of the *S. aureus* sequenced strains MW2, Mu50 and N315 was developed and used to search for user-defined keywords to select genes for probe design. The keywords used for this array were toxin, leukocidin, coagulase, adhesion, protease, nuclease, lipoprotein, lipase, capsular polysaccharide, binding, regulatory and iron. The probes were designed using the OligoArray program (Rouillard *et al., 2002*), a program that computes gene specific and secondary structure free oligonucleotides for genome-scale oligonucleotide microarray

construction. Each probe on the array consisted on 50 bases, with a calculated melting temperature (T_m) of approximately 80°C and with minimal internal structure. Probe specificity was verified through a local BLAST database containing the complete sequences of MW2, Mu50 and N315.

The unannotated genome of *S. aureus* strain MRSA252 (which has since been annotated) was also used to identify additional virulence-associated genes by searching for protein motifs associated with virulence. Potential genes encoding toxins and adhesins were selected using this method. Additional virulence-associated genes that were not present in the genomes of the four sequenced strains (e.g. staphylococcal enterotoxin E), selected from the literature, were also included. Finally, the seven house-keeping genes used in MLST together with the 16S and 23S rRNA subunits were included as positive controls. The complete array included probes for 384 genes.

During this project, the original design of the array has since been extended to include an additional 337 *S. aureus* virulence-associated genes whose sequences have been obtained from more recently sequenced strains such as the USA300 strain. The new genes were selected using the methods described above. New keywords used in the genome browser included proteinase, peptidase and SCC. A further 30 *E. coli* gene probes were also added as negative controls. These provide a system of calculating background hybridisation noise, and therefore enable elimination of noise from real positive results in the data analysis. The revised array comprises probes for a total of 751 genes. Gene descriptions were updated as new information became available; all annotation descriptions here are based on the June-2008 revision (see CD for complete gene list).

2.4.1.2 Printing

2.4.1.2.1 Glass slides

Oligonucleotide probes were printed onto epoxy coated glass slides (Schott, Nexterion E slides). The original probes were 5'-amino-linked, to enable covalent binding of the free amino group to the glass surface during microarray preparation. However, it was found that non-amino linked probes bind to the epoxy surface efficiently with no observable difference in hybridisation so the majority of probes used were of the non-amino linked type.

2.4.1.2.2 Printing platform

2.4.1.2.2.1 Biorobotics Microgrid II

The Biorobotics system is able to print up to 108 slides from up to 20 microtitre plates held in a cooled storage rack in an unattended run. Probes were printed at a concentration of 25pmol/μl in ArrayIT spotting solution (Molecular Solutions Europe Ltd), using tungsten split pins (Biorobotics) at ambient temperature and humidity. Slides were printed with 16 pins to give a 4 by 4 array grid, each block contained 16 rows and 16 columns giving a potential total of 4096 features. Four replicate spots were made for each probe. This system suffered from the generally low ambient humidity in the laboratory which adversely affected spot morphology and caused drying and subsequent blocking of the print tips. Later microarray printing was done on the Genetix Qarray mini.

2.4.1.2.2.2 Qarray mini (Genetix)

The Qarray mini proved efficient and reproducible for array printing. The humidity was controlled at 50%, Stainless steel TeleChem SMP4 split pins (TeleChem) were used. The layout of the arrays was changed to include five replicates, printed at random positions within the grid. In this way, the maximum information about a particular gene can be gained even in the likely event of spatial artefacts post hybridization. This was a problem seen in the previous layout, where sometimes the information for all replicates was lost.

2.4.2 Hybridization, scanning and preliminary data generation

2.4.2.1 Hybridization

Hybridisation (7h) was performed using the Lucidea SlidePro instrument (GE Healthcare). Mixed target samples (Cy3 and Cy5 labelled cDNA or DNA) in GenHyb buffer (100µl; Genetix, UK) were denatured (95°C, 3min) and injected into the hybridization chambers of the Lucidea SlidePro. The targets usually represented approximately 50% of the total eluate from the MinElute clean-up for either DNA or cDNA labelling. Following hybridization, slides were washed sequentially (under non-stringent conditions) with 1X SSC/ 0.2% SDS, 0.1X SSC/ 0·2 % SDS, 0.1X SSC and propan-2-ol then finally air dried.

2.4.2.2 Scanning

In the beginning of the project, arrays were scanned using the Affymetrix™ 428 array confocal microscope scanner. This was later replaced by the Genetix aQuire scanner. Fluorescence intensity for the Genetix aQuire confocal laser scanner was measured

concurrently at 532nm (Cy3) and 639nm (Cy5) using a resolution of 10μM. The images were then analysed using QScan software (Genetix), which superimposes the two images.

2.4.2.3 Image processing

For each array, the process of spot finding begins by positioning the grid (defined by the gal file) which indicates by an array of circular areas the expected size and position of each spot. The grid is placed over the scanned spots and resized if necessary so that the spots fall within the circles as much as possible. Individual circles on the grid may be adjusted manually where necessary. Circle areas may also be adjusted to increase or decrease the number of pixels included in the spot to more clearly define the signal and reduce the background intensities.

2.4.2.4 Generating data statistics

The gene IDs (associated with each spot) were loaded from a formatted list. The mean, mode, median intensities and standard deviations for pixels in both the spots (defined areas) and their associated backgrounds were measured. The data was exported as a CSV file for more detailed analysis in Microsoft Excel.

2.4.3 Further data analysis (Microsoft Excel)

2.4.3.1 Background subtraction

Initially, for each defined spot on the array, the background intensity (mean intensity of pixels within a doughnut shaped area surrounding the spot) was subtracted from signal

intensity (mean intensity of pixels within the circle defining the spot). The ratios of the background-subtracted two colour intensities were also calculated.

Median values were calculated for all replicates of each probe, and henceforth used to represent that gene.

2.4.3.2 Normalization, logarithmic transformation and cut-off determination

Individual probe spot ratios were normalised by dividing by the ratio of the means of all spots so that the ratios clustered around a value of one. This method performs a crude adjustment for the variable specific activities of the target preparations (including different dye incorporation rates) and corrects for different scanner settings used for the two dyes. Signals that were greater than the mean plus four standard deviations of the negative controls (*E. coli* specific probes) were considered positive except when the $\log_2$ ratios of fluorescence values (Cy3/Cy5) were outside of the normal distribution. The cut-off values on either side of the normal distribution of $\log_2$ ratios were selected automatically from the Minimum Kernel Density algorithm. Similarly signals that were greater than the mean plus two standard deviations of the negative controls were considered equivocal.

2.4.3.3 Data filtering

Spots that were defined as 'bad' from the raw image (due to poor morphology or spatial artefacts) were excluded from the analysis. Furthermore, if 4 of the 5 replicates of a gene were defined as 'bad', the result for that gene was considered equivocal. In addition, each probe result for the control strain was checked against the known sequence of MW2 and if found to be incorrect the test strain result was discarded. Using these methods, low quality and/or questionable data were flagged for further analysis or discarded directly.

2.4.3.4 Heat-map generation

Heat maps were generated in Excel. Initially, the data for all the strains were collated into an Excel spreadsheet, where each column contained the full CGH result (positive or negative) of one strain. Positive results were coloured red, negative green and the equivocal remained white (see CD for complete data set in colour). For those results presented in this thesis (unless otherwise stated), positive results are coloured black, negative grey and the equivocal remained white.

2.4.3.5 Cluster analysis

Cluster analysis was performed in BioNumerics™ (Applied Maths), using the simple matching coefficient and UPGMA dendogram type. The heatmap data was converted into binary digits, and saved as a comma-delimited text file, which could then be exported into BioNumerics™ to produce the clusters.

CHAPTER 3.0 *S. AUREUS* MICROARRAY DEVELOPMENT AND EVALUATION

3.1 *S. aureus* microarray development

The methods often used to delineate the diversity and/or relatedness of *S. aureus* strains circulating within both healthcare institutions and community settings include *spa* typing, ribotyping, PFGE, VNTR, MLST and SCC*mec* typing. However, none of these techniques extensively defines the genes that constitute the organism(s) under investigation. This requires further evaluation of the genes of interest within an individual isolate. This is currently achieved by PCR amplification of the particular loci of interest followed by restriction enzyme analysis or sequencing (Dunman *et al.*, 2004).

A microarray was designed to provide an alternative method of characterizing *S. aureus* strains, using array features (probes) to differentiate sub-types while concurrently providing enhanced information concerning the genetic composition of strains. The design also allowed the array to be used as a tool to investigate the natural population dynamics of *S. aureus* strains and to better understand their potential to cause disease.

The earlier stages of this project focussed on method development, both for CGH and transcription profiling. The methods evaluated were detailed extensively in the previous section (chapter 2.0). Here the methods leading to the final protocol, used to generate the results, are presented and discussed.

3.1.1Array Design

3.1.1.1.Array Layout

The original probe layout printed on the BioRobotics instrument included four replicates for each probe spotted in adjacent positions (fig. 3.1a). The limitation of this design was apparent post-hybridisation where local artefacts occasionally affected all replicates for a particular probe. This design layout was improved by printing probes at random positions within each sub-grid (fig. 3.1b). In addition, the complete arrays were also printed in duplicate on each slide, which significantly reduced the loss of data.

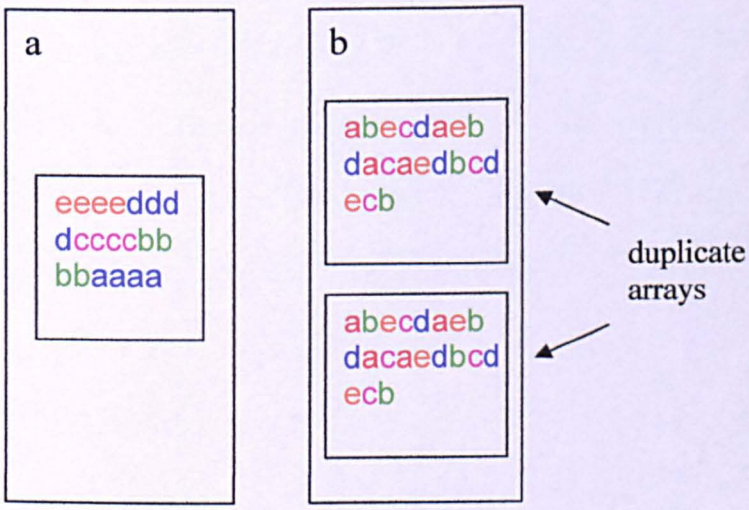


Figure 3.1 Array layout (a) previous design (b) adapted design layout.
In the original array design, replicate probes (represented by letters) were printed side by side. In the adapted array layout, probes were printed in random order.

3.1.1.2 Negative control probes

Escherichia coli probes were spotted onto the arrays for use as negative controls during data analysis. These probes provided a system for calculating background hybridisation

noise, and therefore enabled its subtraction from the signal. This is important because background may be generated by the scanner as well as by non-specific binding. The former is apparent within the statistics collected by the scanner software and can be eliminated later, whilst the latter can be estimated by means of negative controls such as these. This is detailed further in the data analysis section (section 3.1.3).

3.1.1.3 Printing

The order of the source plates (used for printing) was rearranged to ensure printing of the *E. coli* negative controls first, and the ribosomal RNA probes (rRNA) (which generated the most signal in studies of the transcriptosome) last. This arrangement was chosen to prevent possible contamination of the negative controls and other probes by the rRNA, which occurred on one occasion where there was an error in setting up the printing robot. Contamination was only evident in hybridisations to total RNA when, for example, probes printed immediately after the rRNA probes were unexpectedly positive.

3.1.1.4 Probe design

All additional probes (n=337) were ordered as 50mers, to match the primary batch. Previous research conducted on establishing the optimal length for probes have suggested 70mers as the best compromise between specificity and sensitivity. However the array described here includes many genes that belong to families. Consequently, the selection of 70mers with high specificity was problematic. Besides the advantage of high specificity (in the context of this study), 50mers are significantly less costly than 70mers (approximately 25% of the 70mer cost) since cartridge purification is unnecessary.

3.2 Protocol evaluation

The microarray was designed for two main types of study (i) CGH, and (ii) transcription profiling. CGH involves the analysis of samples at the level of DNA, and therefore provides a profile for the presence or absence of genes on the array. Effectively, this highlights genetic differences between the strains under investigation. Transcription profiling, however, consists of analysis at the level of RNA and provides information implicitly about the function of genes through analysis of their pattern of expression or regulation. However, as RNA has a very short half life, it can be converted into more stable cDNA through reverse transcription. Protocols for converting RNA to cDNA that maintain the *in vivo* gene expression profile were developed during this project.

3.2.1 Development of a protocol for *S. aureus* transcription profiling

An important aspect of this project was to establish a cost-effective and reproducible protocol for transcription profiling of *S. aureus* strains. The biggest limitation of transcription profiling protocols is the requirement for relatively large amounts of RNA (20-100μg total RNA) to produce an adequate signal over background noise. This is particularly important for detecting low abundance transcripts. Where the RNA yield is low, such as in clinical samples, there are two possible approaches to reduce the quantity of RNA required; signal amplification and sample amplification.

Signal amplification involves improving the labelling efficiency in order to produce a greater number of signal molecules per transcript. Conventional labelling methods involve the direct incorporation of dye molecules in a reverse transcription reaction. However, this approach suffers from several limitations including a high cost for the dyes and inefficient incorporation of dye molecules due to their bulky structure. Furthermore, the dye

molecules may be incorporated at different rates in different transcripts, a factor that will affect the amount of signal generated from these transcripts. More recently, indirect labelling techniques have been developed to overcome some of the labelling issues of the direct method. In this project, indirect labelling via aa-dUTP and aha-dUTP were tested (section 2.2.7.4). The amino-allyl and aha groups are small compared with the standard bulky fluorescent dyes and have a minimal impact on incorporation of the nucleotide in a reverse transcription reaction.

The second strategy, sample amplification, involves global amplification of the sample. Here the most important issue is the conservation of relative transcript abundance. In recent years, many methods for RNA amplification have been developed. These have included *in vitro* transcription using the Eberwine protocol, single primer amplification, SMART technology (Clontech), Ribo-SPIA™ RNA amplification (NuGEN), amplification using terminal continuation, and others. However, the majority of protocols are tailored towards the amplification of eukaryotic RNA by exploiting their characteristic poly-A tail sequences. The limited availability of a protocol targeting prokaryotic sources necessitated the pursuit of an in-house sample amplification protocol. This section details protocols developed to investigate the transcription profiles of *S. aureus* strains.

3.2.1.1 RNA isolation and concentration: optimising the RNA yield

3.2.1.1.1 Optimising the RNeasy protocol

The primary method used for the isolation of RNA from cells grown on solid and liquid agar was the Qiagen RNeasy mini kit. Good RNA yields (0.2µg/µl) were reproducible after optimization. Most significantly, the initial RNA stabilisation step implemented prior to RNA extraction improved the RNA yield considerably (figure 3.2). RNA stabilisation

was initially performed using RNAlater (Qiagen), but later exchanged with RNAprotect (Qiagen) which proved more efficient as it is targeted towards use in bacterial samples. RNAprotect stabilises RNA prior to cell lysis in order to maintain the gene expression profile during the extraction process. These RNA stabilisation products (Ambion patent, 6,528,641) are buffered salt solutions (e.g. ammonium sulphate) that act by rapidly salting out nucleases present within target cells.

In the early stages of the project, the RNeasy kit provided a yield of 0.1μg/μl in comparison to the maximum 1μg/μl quoted by the manufacturer. As a means of increasing the RNA yield, various methods of optimizing the protocol were explored. These included using a larger number of cells for RNA extraction, incorporating SDS to break any RNA-protein bonds that may potentially prevent the RNA from sticking to (or eluting from) the Qiagen column membrane, growing cells in liquid culture instead of on solid agar (since actively growing cells contain larger quantities of mRNA) and finally optimizing the RNA elution step (discussed below). The 16S primers were used to detect the presence of chromosomal DNA remaining post- DNase treatment during the RNA isolation procedure. DNA-free-RNA samples are essential during subsequent cDNA synthesis to ensure the results on the array are representative of gene expression and not contaminating chromosomal DNA. In particular, since the array was designed to detect both DNA and cDNA targets, there would be no way of differentiating these post-hybridisation based on signal alone.

Figure 3.2 shows RNA extracts (using the RNeasy kit) taken from strain NCTC 8325, obtained using different amount of starting material, pre- and post-DNase treatment. Cells were extracted from plate cultures. The gel image shows increasing amounts of harvested cells (except lane 2 which was part of another test). These same preparations were DNase treated and shown also in the figure. In general, the RNA eluted post DNase treatment

produced roughly the same yield. Presumably the binding capacity of the clean-up kit played some factor in this. However, lane 7 (which contained 2µl cells) produced the largest RNA yield post-DNase treatment (lane 8). Although this particular preparation appears degraded in the gel image (a problem experienced initially), it nevertheless showed that that using optimised cell numbers was effective in obtaining higher RNA yields. Consequently, larger cell numbers (than the initial protocol) were used for harvesting cells from plate cultures. The limitation of this method was that it was not reproducible since the exact number of cells used was unknown. This could have been overcome by resuspending cells in solution and measuring the OD of the suspension, but this method was replaced with quantified liquid cultures (described below).

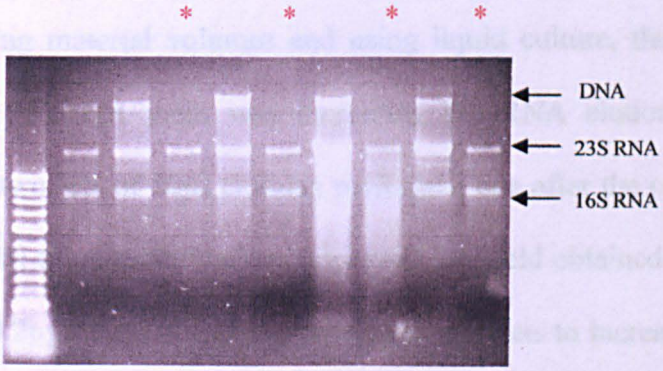


Figure 3.2 RNA extractions from strain NCTC 8325-4; evaluating the optimal size for cells harvested from plate culture. Lane 1, DNA Hyperladder II; lane 2, part of a separate experiment; lane 3, 0.5µl cell harvest; lane 4, DNase treated preparation of lane 3; lane 5, 0.75µl cell harvest; lane 6, DNase treated preparation of lane 5; lane 7, 2µl cell harvest; lane 8, DNase treated preparation of lane 7; lane 9, 5µl cell harvest; lane 10, DNase treated preparation of lane 9. DNase digested preparations are marked (*).

A possible reason for the sub-optimal RNA yields was that RNA was either not sticking to the column membrane or it was not eluting efficiently, possibly due to the formation of RNA-protein complexes. SDS, which tends to disassociate RNA-protein complexes, was included in the lysis buffer but did not significantly improve the RNA yield. It was

therefore concluded that RNA-protein complexes were not a persistent problem, and the SDS step omitted from the final protocol.

A further variation of the original protocol was the extraction of RNA from exponentially growing cells in liquid culture. This was consistent in providing high yields of RNA. It was hypothesised that this was due to the extraction of messenger RNA (mRNA) that is present in greater amounts during exponential growth than in stationary phase (as is the case with overnight plate cultures). Therefore, RNA was extracted from liquid culture, quantified and stabilized in RNAprotect. The RNeasy instruction recommendation of 1 x 10^9 cells as an optimal amount of starting material proved sufficient.

Aside from the starting material volumes and using liquid culture, the most significant factor in improving the RNA yield was increasing the RNA elution volume to two successive steps (i.e. two lots of 50µl elutions performed one after the other), instead to a single 50µl elution. All these modifications improved the yield obtained from the RNeasy kit from 0.1µg/µl to 0.2µg/µl. At one point, pooling cell pellets to increase the RNA yield was evaluated, but this did not provide a significant improvement. The limitation here was the binding capacity of the column membrane used for extraction and purification.

3.2.1.1.2 Comparison of different RNA isolation kits

The RNA yield obtained from the RNeasy kit did not reach the manufacturer's quoted yield even with the protocol adaptations described above. To achieve higher yields, a number of different kits (both small- and large-scale) were evaluated for the quality and quantity of RNA produced. Kits were obtained from several suppliers (Qiagen, Promega, Invitrogen and GE Healthcare) and compared with the RNeasy mini kit.

In this study the least efficient kit was the Promega PureYield kit, which gave a very low RNA yield (in a single test, <100ng was obtained at a concentration below 20ng/µl). The GE Healthcare kit gave RNA depleted of small RNA when analysed on an agarose gel (not shown). Although this is favourable towards concentrating RNA, it was unclear whether other RNA transcripts were also being lost. This kit was therefore not pursued.

Of the large scale extraction kits, the Qiagen RNeasy midi kit was the most successful. However, although providing a good yield (40µg), the concentration was similar to that obtained from the RNeasy mini kit. Furthermore, the midi kit was more time consuming and labour intensive than the mini kit, a limitation when extracting from many samples. It was therefore not pursued for these reasons. It was concluded that none of the alternative kits provided consistent advantages over the RNeasy mini kit used with the modified protocol and this kit was retained as the standard method for RNA extraction.

3.2.1.1.3 RNA concentration

Methods for concentrating RNA were also investigated. This was important because the subsequent cDNA synthesis reactions were limited to a maximum sample input volume of 20µl containing at least 5µg RNA. Standard RNA precipitation and vacuum evaporation (i.e. Spin-Vac) concentration methods were evaluated. These allowed testing of input RNA concentration of 0.8µg/µl (50µl reaction mix) in the cDNA synthesis protocols. However, concentrated RNA gave poor results in the transcription reactions as judged by poor signal intensities on the post-hybridisation array images. Furthermore the vacuum concentration method resulted in RNA degradation as judged by gel analysis (not shown). For these reasons, these methods were not included in the routine protocol.

An alternative method for RNA concentration was the Ribominus kit (Invitrogen) which removes rRNA from total RNA. Since rRNA competes with mRNA in reactions dependent on random primers it was thought that the presence of large amounts of rRNA might reduce mRNA transcription and labelling significantly. Therefore, reducing the levels of rRNA should result in more mRNA transcripts being produced. However, gel analysis of RNA post Ribominus treatment (not shown) failed to show any material on the gel image indicating that no (or very little) mRNA was extracted from cells, or that both the mRNA and rRNA were being removed by the Ribominus method or finally that the mRNA was not recovered from the binding column used in the kit.

3.2.1.1.4 RNA quantification

In the initial stages of the project, RNA integrity was determined by standard gel electrophoresis analysis and the yield estimated by comparison to molecular standards on the gel. Denaturing gel electrophoresis did not provide a significant improvement in the analysis of RNA, and required more toxic reagents. It was therefore not used routinely. In later stages, the Qubit quantification system (Invitrogen) was used post-extraction to provide a more accurate and faster method of quantifying the RNA yield.

3.2.1.2 Growth curves

The growth kinetics of the test isolates were monitored by taking aliquots and measuring OD_{600} readings and performing viable counts. The objective of this work was to monitor the time points of growth to estimate lag-, log- and stationary-phases of growth. Cells extracted from log- and stationary-phases would be subsequently used for transcription profiling. The logarithmic phase is assumed to be the stage where the greatest amounts of mRNA is being produced, reflecting the transcription of genes necessary for growth and

other metabolic processes. At the stationary phase, the rates of cell-division and of many other processes is reduced although expression of key genes associated with survival under adverse conditions is likely to be enhanced.

The growth patterns for the *agr*$^+$ strain consisted of a short lag phase, followed by an exponential phase, then a period of stable cell numbers (possibly even a small decline) before the cells resumed growth. In addition, the data suggest that the cells were still growing after 23h (fig. 3.3).

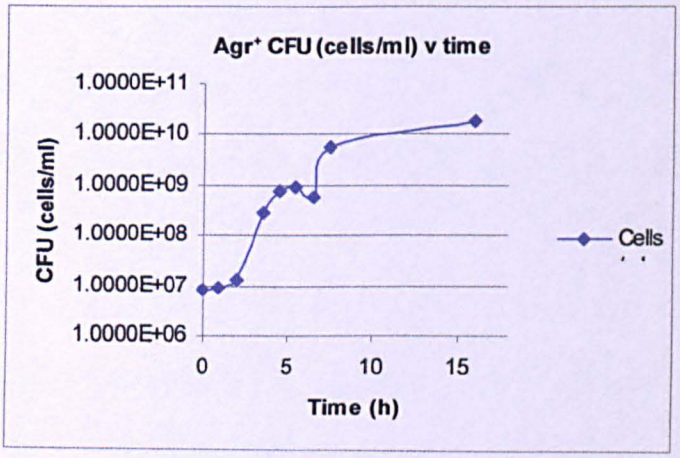


Figure 3.3 *Agr*$^+$ growth kinetics over a 23h period.

The growth curve for the *agr*$^-$ strain was very similar to that of its isogenic partner except that the period of stability following exponential growth was less clearly defined (fig. 3.4).

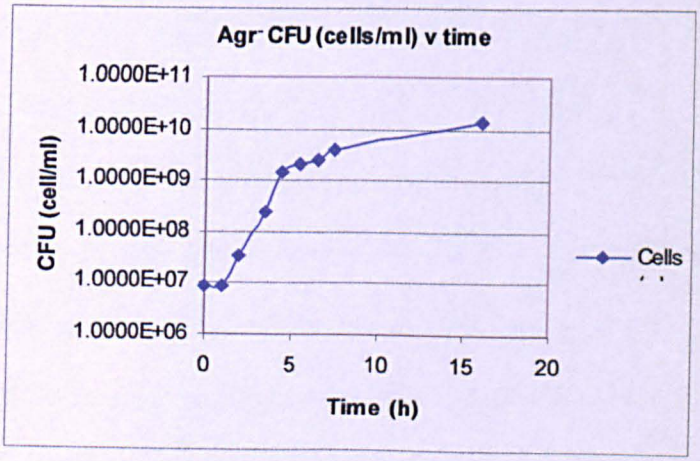


Figure 3.4 *Agr*$^-$ growth kinetics over a 23h period.

From these data, and similar work (Sabersheikh, 2003), it seems that *S. aureus* rarely display the classic sigmoidal growth curve and often show more than one phase of rapid growth. It seems likely that the rapid depletion of glucose and other readily available nutrients occurs during the exponential growth phase and that this is followed by a period of re-adjustment followed by sustained non-exponential growth over a longer period during which there is greater competition for the available resources.

For the transcription profiling study of the *agr* mutant strains, the expression profiles for early-log (5h), late-log (7.5h) and overnight cells (23h) were analysed. As expected log phase gave the most heterogeneous mRNA (most spots seen on the post hybridisation microarray images). The overnight cultured cells were transcriptionally active but fewer transcripts were found at detectable levels. The results of this study are detailed in chapter 5.0.

3.2.1.3 cDNA synthesis: signal amplification and sample amplification

Several parameters of standard cDNA synthesis protocols were evaluated. These methods may also include amplification and labelling steps (both direct and indirect). Primarily, since prokaryotic mRNAs lack polyA of sufficient length for efficient oligodT priming, only random priming was viable in these protocols. The biggest challenge observed during cDNA synthesis was the production of the second-strand in the pursuit of indirectly labelled cDNA. It appears that the indirect label incorporated into the first cDNA strand may have interfered with the production of the second cDNA strand. Consequently, reverse transcriptase variants were tested with these protocols to overcome any steric effects that may prevent any particular enzyme transcribing the modified nucleotides incorporated into the cDNA.

Initially, the Bioprime kit (Invitrogen) for the direct labelling of DNA/cDNA was evaluated. This method, which depends upon random primers and direct incorporation of Cy-dye labelled dUTP by reverse transcriptase, worked well for some samples but was not sufficiently robust to provide good fluorescent signals when the input RNA was of lower quality. Since good signals are essential for accurate transcript quantification, alternative protocols were sought. The problem of inefficient Cy-dye labelled nucleotide incorporation has been recognized by the commercial market and alternative and cheaper indirect labelling methods have been developed using amine-modified dNTPs (section 2.3.4.2). The smaller structure of these modified nucleotides (compared with the relatively large Cy groups) enables them to be incorporated into the target nucleic acid at a rate similar to that of unmodified nucleotides. The Cy-dye (or Alexa-dye) can then be bound to the target DNA post-synthesis by ester bonding.

In order to overcome the problems associated with limited mRNA yields, experiments were performed to adapt PCR for generic amplification of cDNA. Initially, first strand cDNA synthesis using commercial Random Primer 9 (NEB) was compared to that obtained with an in-house primer comprised of a random nonamer with a cytomegalovirus (CMV) specific 5'-tail. Custom synthesised oligonucleotides are less expensive than off-the-shelf nonamers (Random Primer 9). For the second strand synthesis, a random nonamer with a complementary CMV primer tail was used. The use of CMV-tailed random nonamers (CMV-forward and CMV-reverse) for the first and the second strands of cDNA synthesis was designed to generate template cDNAs that should be efficiently amplified using non-tailed CMV primers. A previously developed RT-PCR using the CMV primers (Kearns *et al.*, 2002) showed their reproducibility in detecting their targets. The melt curves, used to confirm PCR product identity, have been shown to be reproducible at 68°C. For this reason, these primers were incorporated into cDNA synthesis via the tailed nonamers.

The reverse transcriptase enzyme MMuLV was used with the tailed nonamers to incorporate aha-dUTP into first strand cDNA, an adaptation of the DeRisi cDNA synthesis and indirect-labelling protocol. During second strand synthesis, two protocol variants using either exo- Klenow or Taq polymerase were evaluated. Priming with the untailed primer was compared with that achieved using the tailed primers. The cDNA produced was analysed by real-time PCR (using either the primers complementary to the CMV tail, or SYBR Green). It was shown that cDNA amplification occurred and different yields were obtained when different enzymes were used to produce the cDNA i.e. *Taq* polymerase and exo- Klenow fragment (data not shown). The yields were generally low, indicating these protocol variants were relatively inefficient (data not shown). However, the tailed nonamers were found to be slightly more efficient at producing intact cDNA (data not shown). Successful cDNA synthesis was confirmed by a logarithmic amplification curve and by a defined melt peak at 68°C. In all cases with the different enzymes, post-hybridisation signals were low, indicating inefficient incorporation. The possibility remains that this was due to inadequate RNA extraction and labelling. However, gel analysis of the RNA showed relatively good yields and direct labelling protocols showed more spots on the array. The results of the different protocols were as follows (data not shown).

The PCR methods were not found to be reproducible with either aha-dUTP or amino-allyl dUTP (used in the original DeRisi protocol). Optimisation of the aha-dUTP:dTTP ratio and using a lower aha-dUTP concentration were found to give higher yields of product as judged by crossing thresholds during PCR analysis. Yet, by reducing the aha-dUTP concentration, fewer Cy-dye molecules would subsequently label to the cDNA, confounding the purpose for pursuing indirect labelling techniques. Direct dye

incorporation of dUTPs modified with Cy3, Cy5, or the alternatives Alexa-555 and Alexa-647 were tested, but these materials also adversely affected the PCR.

As an alternative to PCR amplification, whole genome amplification methods (originally designed for the amplification of DNA) were tested for their ability to produce and amplify cDNA. The first protocol tested used random nonamers (N^9) for first strand cDNA synthesis followed by whole genome amplification using the GenomiPhi (GPhi) kit (GE Healthcare). GPhi was used to produce the second cDNA strand, whilst at the same time amplifying the two strands (under isothermal conditions). Real-time (RT-) PCR was employed for analysis of the products. Primers specific for 16S RNA sequences were used in the RT-PCR. Successful whole genome amplification should produce a large number of copies of each component sequence recognised by a small number of cycles before the crossing threshold is reached. However, the PCR showed that the reaction products of this protocol contained low cDNA levels. Examination of the product of whole genome amplification revealed that non-specific amplification had occurred since all negative controls showed a high molecular size amplification product. This phenomenon has been acknowledged by the manufacturer and a second version of the GPhi enzyme is now available.

An alternative in-house whole genome amplification protocol similar to the GPhi kit (which utilises Phi29 enzyme) was evaluated so that Phi29 from an alternative source (NEB) could be tested. It was hoped that the NEB enzyme would contain lower levels of contaminating DNA. This in-house method gave higher yields of high molecular weight DNA compared with the GPhi kit but unfortunately was not efficient at amplifying the relatively short cDNA sequences that were the intended targets in the reaction (revealed by RT-PCR). Generally, it was noted that the use of whole genome amplification with cDNA targets did not give significant amplification of the short sequences. One possibility for this

low finding is that the whole genome amplification process is relatively inefficient for short sequences (i.e. cDNA) compared with long sequences (i.e. chromosomal DNA). Concatenation of cDNA may be a solution to this problem but this was not evaluated in this study.

Gel analysis of cDNA generally showed the presence of residual rRNA (not shown). This indicated that the RNaseH hydrolysis was not efficient. RNaseH degrades RNA in RNA:DNA heteroduplex complexes. Where cDNA synthesis was inefficient, less heteroduplexes will have formed and therefore less RNA will have been degraded. To remove residual RNA, alkaline hydrolysis (NaOH) and final neutralisation (HCl) steps were added to the protocol resulting in the production of cDNA containing no RNA.

The inefficiency of the second-strand synthesis and amplification methods when applied to *S. aureus* mRNA greatly reduces the potential advantages of using these methods. Furthermore, the addition of manipulations increases the possibility that the final cDNA product will not be representative of the cellular mRNA. Consequently, it was decided that these steps would not be used in subsequent studies. First strand cDNA indirectly labelled with the fluorescent dyes by random priming can generally give reproducible signals in microarray work. Additionally, random priming by some reverse transcriptases (with DNA properties) can produce the second cDNA strand during the first strand synthesis, the evidence for which can be observed post-hybridisation. That is, the design of the oligonucleotides of the virulence-associated microarray was such that they were synthesised to capture cDNA in opposite reading orientations (3'-5' and 5'-3'), as is the case for the two strands of cDNA. It was observed that the transcripts produced using the reverse transcriptase MMuLV for first strand synthesis produced signals in some of the oligonucleotides designed to recognise the second strand (data not shown). However, the intensity of most of the probes remained weak. The ChipShot kit (Promega) was therefore

used in subsequent studies. This kit employs amino-allyl dUTP incorporation via a reverse transcriptase (unspecified). The ChipShot kit successfully and reproducibly produce indirectly labelled cDNA from *S. aureus* RNA. The only limitation of this method was the requirement of at least 2.5µg of total RNA to produce a high quality array image. This was achievable with the optimised RNeasy protocol.

3.2.2 Development of a protocol for comparative genome hybridisation

3.2.2.1 DNA extraction

DNA extraction with the Qiagen DNeasy kit was optimised. The manufacturer's recommended lysis conditions were sub-optimal and the lysis step described for RNA extraction (section 2.2.4.1) was used. During protocol evaluation, the Pitcher method (Pitcher and Saunders, 1989) was also evaluated but this was not as convenient as the Qiagen kit and the purification of DNA from RNA was less efficient. The DNeasy kit was therefore used routinely for DNA extraction.

3.2.2.2 DNA labelling: signal amplification and sample amplification

The attempts to produce indirectly-labelled cDNA led to the development of a protocol for indirect labelling of DNA via aminohexyl tagged (from aha dUTP) intermediates. Primarily, a commercial kit (Bioprime kit) was used. The kit generated good quality labelled DNA and spots of good intensity (bright signal) were seen on the post-hybridisation images.

The quality of microarray data generated is determined by the quantity and label density of the target DNA. These, in turn, are determined by the quality of the DNA extract, the ratio

of modified to native nucleotides and the efficiency of coupling with the NHS-ester. However, during optimisation of the labelling protocol it was noted that the colour intensity (absorbance) of the product eluted from the MinElute clean–up column following the NHS-ester reaction was strongly correlated with microarray signal intensity. That is, a brightly coloured sample (prior to hybridisation) would give a strong signal on the array, whereas a sample with no colour (or with a slight tint) did not produce a good quality array signal. Thus colour intensity was used as surrogate for effective labelling.

The kit protocol has been optimised for a defined quantity of DNA. However, due to inter-strain variation, DNA extracts were occasionally of low yield. Furthermore, it was considered desirable to be able to store an archive of DNA preparations, therefore, the whole genome amplification (WGA) method was evaluated. The GPhi kit (described previously, section 2.3.3.1) was tested with chromosomal DNA, and produced amplified DNA of large molecular weight. However, the negative controls always showed an amplified product (as discussed above). The NEB Phi29 protocol also suffered from similar limitations. It was considered likely that the DNA amplified in the negative controls was of bacterial origin (i.e. a contaminant of the Phi29 enzyme) and consequently that it may influence array results. Thus, WGA was not used to generate DNA for use in array studies. Instead in order to overcome the occasional problem of low DNA concentrations from the DNeasy columns, diluted solutions were dried on a heating block set at 65°C and resuspended in a suitable volume of water. This simple yet effective method overcame the possible contamination, time and cost associated with DNA amplification techniques.

The protocol adopted for DNA labelling was an adaptation of the DeRisi cDNA labelling protocol (http://cmgm.stanford.edu/pbrown/protocols/amino-allyl.htm). The adaptations made were as follows. The random primer was a cocktail (equal weights) of 8, 9 and

10mers (rather than 9mers). Furthermore, a single dye pack (intended for one sample) was used to label three samples without any apparent reduction in the labelling efficiency of the DNA. This measure significantly reduced the cost of performing microarray experiments. Finally, the MinElute cleanup kit (optimised in the DeRisi protocol) worked efficiently for intermediate and final cleanup steps in the protocol.

3.3 Microarray analysis and data mining

3.3.1 Pre-analysis

For each spot on the array, the mean intensities and standard deviations for pixels in both the spots and their surrounding local background areas were measured. The local background and non-specific background (*E. coli* mean values) were subtracted from the mean value of each spot. Several tests of the data were performed to distinguish between positive signals, low signals representing cross-reactivity and background noise. Abnormal spots either marked as 'bad' manually (prior to quantification) or automatically by discarding spots with high pixel to pixel variation were excluded from the analysis. If all four replicates of a spot were marked as 'bad' spots, or if only one good value remained, the result was highlighted and classified equivocal. The ratio between the mean (or median, see below) signal intensities in the red (Cy3) and green (Cy5) channels were then calculated for the remaining replicate spots. Individual spot ratios were normalised by dividing by the ratio of the means (or medians) of all spots so that the ratios clustered around a value of one. This method performs a crude adjustment for the variable specific activities of the target preparations (including different dye incorporation rates) and corrects for different scanner settings used for the two dyes.

3.3.2 Further analysis

3.3.2.1 CGH data

For the CGH arrays, data were analysed relatively conservatively. A reaction was only considered positive if, following the elimination of poor spots, there was a minimum of two replicates with signals above background. Hybridisation results were categorized as positive or negative based on the mean or median spot intensities. Signals that were greater than the median plus four standard deviations of the negative controls (*E. coli* specific probes) were considered positive when the fluorescence ratio was in the range approximately $\log_2 \pm 2$. When the ratio was outside this range, the target giving the lower signal was considered negative. The purpose of this correction was to eliminate hybridisation reactions from similar but non-identical sequences. The precise range of ratios was determined from the position of the minima on either side of the major peak around zero on a Minimum Kernel density plot as described by Carter and colleagues (2008) (minima were usually at approximately +2 and -2 on the $\log_2$ scale). Intensity ratios for probes reacting with both strains (i.e. strains being tested on any individual array) were always in the range $\log_2 \pm 2$. These manipulations of the data were performed automatically in Microsoft excel using a workbook formatted for the microarray so that it was only necessary to paste the raw data and run the Kernel density module. Ratio cut-offs determined for each array using the kernel density plots provided a robust and dynamic way of identifying cross-reactions. By avoiding a fixed cut-off, genes were more likely to be classified correctly under the conditions of each particular experiment, therefore reducing the number of false-positives.

As the mean for replicate spots may be skewed by any abnormally high fluorescence values (e.g. dust fluorescence), the medians and an estimate of the standard deviation based

on the median absolute deviation (MAD) were also used as criterions to determine significance. Median values were calculated from all valid replicates of each probe. The advantage of using a method based on medians is that outliers were automatically eliminated. Furthermore, since the final results obtained are governed by the stringency of the analysis, and in particular, the cut-off threshold, genes which hybridise poorly may drift in and out of significance. It is therefore important to determine the threshold carefully. Signals that were greater than the median plus two standard deviations were considered equivocal, and those greater than the median plus four standard deviations were considered positive.

Finally, the data for each array feature was designated positive, negative, weak positive or equivocal for both the experimental strain and the control strain, MW2. The results obtained for the MW2 strain could be validated against the published genome sequence (Baba *et al.*, 2002), and a consensus result generated for all MW2 replicates. Using this approach, discrepancies between individual MW2 probe results with the consensus result for that probe highlighted a possible erroneous test result, i.e. results were partially corrected by reference to the MW2 data. In this way, the test strain result was flagged for further analysis or discarded. The final data set was used for determining similarities/ differences between isolates and for heat map generation (chapter 4.0).

3.3.2.2 Transcription profiling data

Data analysis for the transcription profiling experiments was similar to the CGH data analysis. However, the output results were quantitative (to indicate fold change in expression) rather than qualitative (gene presence or absence). Briefly, as with the CGH experiments, background subtraction (using the *E. coli* probes) and crude normalisation (using the ratio of the mean of all spots) were performed. Scatter plots, in the form of M

versus A (MvA) plots were used as a rapid, visual indication of variation within data. MvA plots were applied to the data before and after any pre-processing procedure (background subtraction, normalisation, etc). These plots show the variability of the data (M [minus] = $\log_2$ R - $\log_2$ G, where R and G represent the red and green dye channels) as a function of the mean (A [average] = ½ ($\log_2$ R + $\log_2$ G)). Figure 3.5 shows an MvA plot for one of the arrays used to generate the *agr* results (section 5.1) before and after normalisation. As with most two colour arrays, the raw data generally show dye bias at high and low signal intensities, which is partially corrected by normalisation techniques. Normalisation reduces this variability by centering the data. Following normalisation, signals greater than the median plus three MAD standard deviations of the negative control probes (*E. coli*) were considered to be significant when the fluorescence ratio was greater than $\log_2$ +2 (up-regulated) or $\log_2$ -2 (down-regulated). Additionally, total probe signal intensity had to be greater than 900 fluorescence units, indicating sufficient signal was detected in both channels. Outside this range, no change in expression was considered. Genes were ranked according to fold change in expression.

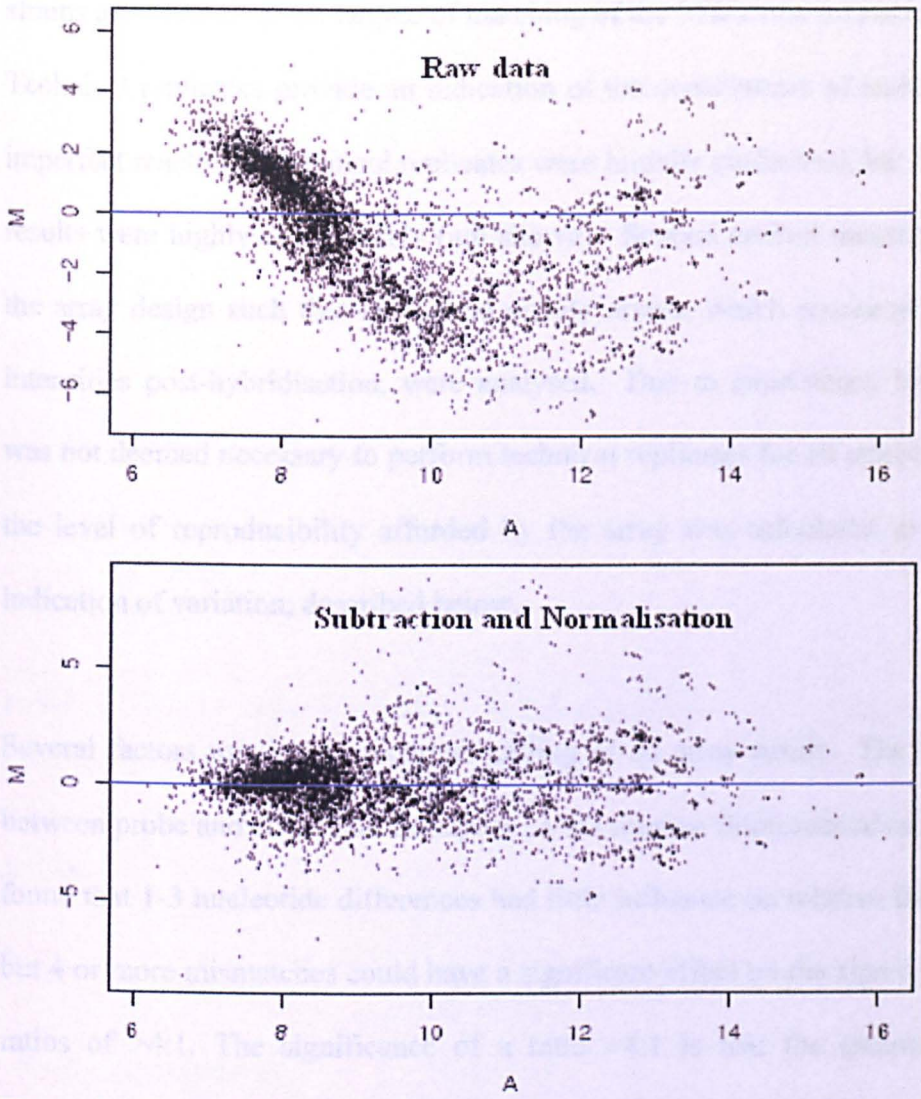


Figure 3.5 Typical MvA plot for array data. The top panel displays the raw data before any pre-processing. The bottom panel shows the spread of the data following background subtraction and normalisation.

3.4 *In silico* validation for selected probes: confirmation of microarray results

3.4.1 Array reproducibility

3.4.1.1 CGH data set

The data for all strains used in the CGH study were collated on an excel spreadsheet. The reproducibility of the array was assessed in two ways. First by comparing replicates of test

strains and second by the degree of matching of the MW2 control strain replicates (n=61). Technical replicates provide an indication of the contribution of technical variation in an imperfect reaction. Technical replicates were initially performed, but it was found that the results were highly reproducible (not shown). Several control measures were included in the array design such that only good quality arrays, which produced bright fluorescence intensities post-hybridisation, were analysed. Due to consistency between replicates, it was not deemed necessary to perform technical replicates for all strains analysed. Instead, the level of reproducibility afforded by the array was calculated to provide an overall indication of variation, described below.

Several factors may cause incorrect calling of an array result. The degree of matching between probe and target has an impact on the relative fluorescence ratio. In general it was found that 1-3 nucleotide differences had little influence on relative fluorescence intensity but 4 or more mismatches could have a significant effect on the signal leading to observed ratios of >4:1. The significance of a ratio >4:1 is that the sample giving the lower fluorescence is automatically called as negative even when the signal is above the cut-off (*E. coli* probe median + two (weak positive) or four standard deviations). Results may also vary for probe/target pairs that routinely give low signals or when probes are missing due to printing errors. To reduce the occurrence of errors, the called data were automatically screened to eliminate results where the control strain (MW2) gave a non-consensus (i.e. incorrect) result. This process results in a loss of information but eliminates poor quality data.

A subset (n=80) of probes were chosen to calculate reproducibility and error rates. Sixty probes were originally designed using MW2 gene sequences and the remainder from MRSA252 genes. The underlying assumptions in these analyses were as follows. The probes designed to hybridise to MW2 genes were a useful set since they were all expected

to hybridise to the control strain (MW2). However, the estimate of reproducibility based on these reactions does not include the effect of variability due to an imperfect match. The MRSA252 genome represents the most divergent genome of the sequenced *S. aureus* strains compared with MW2. Furthermore, the probes from this strain were selected because a corresponding gene was not identified in strain MW2. It is therefore assumed that many of these genes will not have homologues to MW2 (or will have imperfect matches that may not result in hybridisation). Probes from MRSA252 should therefore show some variability in their reactions with MW2. Table 3.1 displays the error rates determined from this data.

Table 3.1 Array reproducibility. *In silico* prediction versus experimental array result based on a subset of probes (n=80) from the MW2 and MRSA252 genome. Part A gives the results for the complete subset of probes, whilst part B shows how these results were derived.

A

		***In silico* prediction**	
		+	-
Experimental array result	+	3842 (82.37%)	18 (0.38%)
	-	37 (0.79%)	767 (16.45%)

B

	Correct negative	**False positive**	**Correct positive**	**False negative**
MRSA252 probes n=20	767	18	266	23
MW2 probes n=60	0	0	3576	14
Total	767	18	3842	37

These results show a 98.82% (true positive + true negative) correct calling rate for the experimental results based on the *in silico* prediction. As discussed above, the MW2 probes could only give true positive or false negative results. The results obtained from the MRSA252 probes, however, could provide estimates of true negative and false positive

rates. Analysis showed that the vast majority of MW2 reactions to probes from the MRSA252 genome, classified as false positive (n=18), were due to low levels reactions between the probes and a similar sequence in MW2 (fig. 3.6). As discussed, a negative result from the test strain coupled with a low but significant level of fluorescence due to the similar MW2 sequence will result in a positive call, whilst positive hybridisation by the test strain leads to the low level MW2 signal being called as negative. Most of the false positives were derived from four probes encoding exotoxins which have somewhat conserved sequences. With respect to the false negatives, the 23 recorded results were derived from only 2 probes (E16-0472, n=13; E16-1208c, n=10) indicating poor probe design since the results were not reproducible due to low signal levels. The true positives recorded between MW2 and the MRSA252 probes were confirmed for 5 probes. In these cases, *in silico* analysis showed matching between the probes and similar sequences (small number of mismatches, fig 3.7) in MW2. However, often the matching was not observed over the complete probe, i.e. sequences either side of the probe were not homologous in strains other than the MRSA252 genome. Finally, the true negatives recorded (n=767) reflect the diversity of the MRSA252 genome compared with MW2.

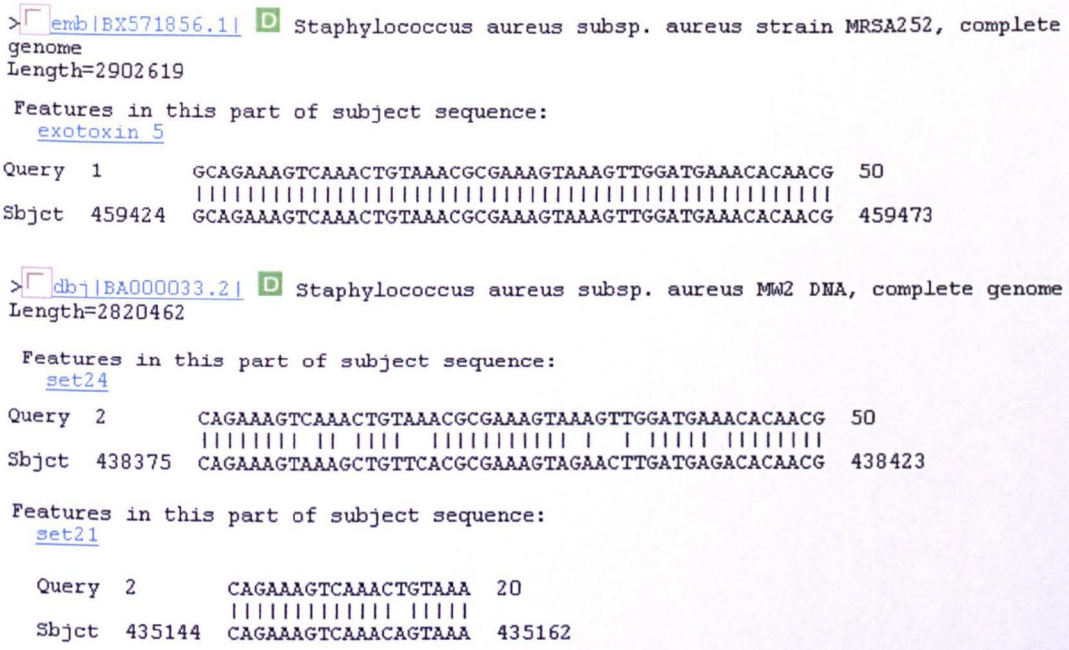


Figure 3.6 BLAST output displaying low level sequence homology between probe E16-0469 (encoding exotoxin 5) and similar sequences in the MW2 genomes, causing false positive probe results. The top panel shows the expected perfect matching with genome sequence MRSA252, from which the probe was designed. The bottom panel displays the sequence homology of this same probe with the MW2 genome, indicating several areas of sequence dissimilarity (e.g. set24) or homology over only a small section of the probe (e.g. set21).

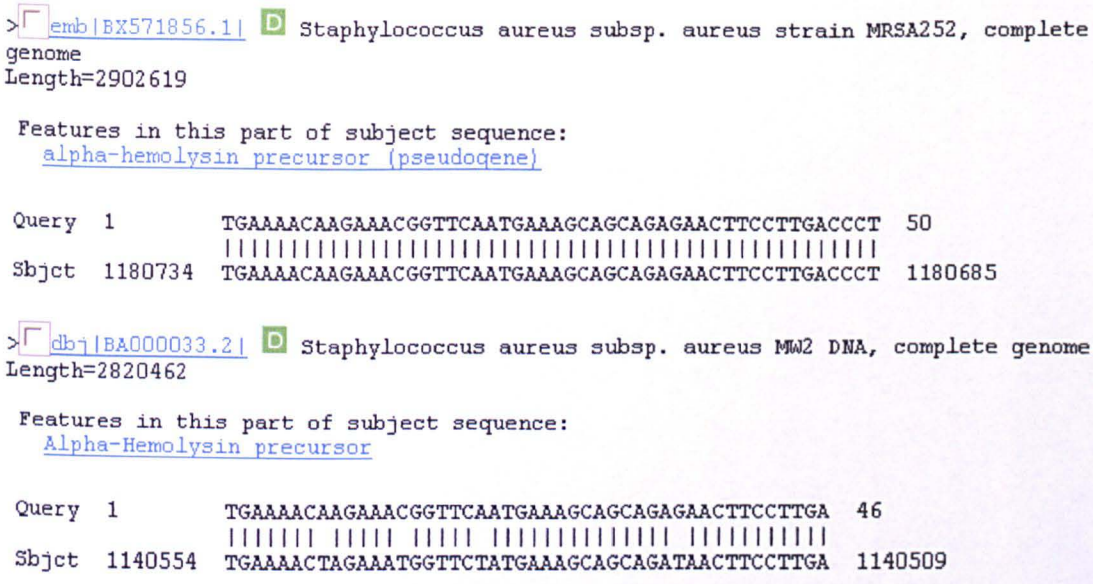


Figure 3.7 BLAST output displaying sequence homology between probe E16-1208c (encoding an alpha haemolysin precursor) and similar sequences in the MW2 genomes causing true positive probe results. The top panel shows the expected perfect matching with genome sequence MRSA252, from which the probe was designed. The bottom panel displays the sequence homology of this same probe with the MW2 genome, indicating overall sequence similarity except for a few mismatches within the probe sequence.

3.4.1.2 Transcription profiling data set

In the transcription profiling experiments, it was not expected that the same expression readings could be generated through repeat experimentation. Instead, where replicates were performed, it was expected that the list of significant genes in each case should be very similar in terms of the genes that are expressed. Yet, the overall degree of up or down regulation should be similar, and therefore if the genes are ordered numerically, the position of genes in this ordered list should be very similar in replicate experiments.

CHAPTER 4.0 COMPARATIVE GENOME HYBRIDISATION

Comparative genome hybridisation (CGH) using the partial composite microarray was exploited to study variation within and between lineages of *S. aureus* to gain insight into their diversity and evolution. Further objectives were to identify strain-specific markers or those with possible linkage to pathogenicity or epidemicity, both useful for epidemiological studies. This chapter has been subdivided to address these aims.

The strain collection for this study comprised a panel of internationally recognised healthcare-associated (HA-) and community-associated (CA-) strains, representative of the most clinically significant lineages in humans. The dominant lineages prevalent in hospital environments belong to CC5, CC8, CC22, CC30 and CC45, whilst those in the community setting include CC1, CC8, CC59, CC80, CC88, ST93 and CC97. Five of the sequenced MRSA strains (MW2, Mu50, N315, COL and SA252) and two of the sequenced MSSA strains (MSSA-476 and NCTC-8325) were also analysed. The results are presented and discussed below (see CD for complete data set).

4.1 Studying variation amongst lineages of *S. aureus*

4.1.1 Clonality of *S. aureus*

The clonality of *S. aureus* has long been recognised through traditional typing techniques. Feil and colleagues (2004) showed this by performing MLST on clinical *S. aureus* strains. International standardisation of typing and nomenclature was greatly improved through MLST, which groups strains into clonal complexes (CC) according to the nucleotide sequences of a core set of conserved genes. Despite the fact the design of the array

comprises the variable and core-variable genes rather than the standard 'house-keeping' genes used for MLST, the data presented here show congruence with MLST (see Table 4.1 and Figure 4.1). An advantage of the array approach is the additional information afforded by this technology (discussed previously, section 1.4). Presentation of the data in heat map format provides a way of analysing the data visually and rapidly. Individual strain profiles can be used as a means of discriminating between related strains. The array provides a valuable and comprehensive tool based on the determination of virulence-associated markers of strains of interest. The data generated can be used as a standard reference when analysing further clinical strains on the array.

As well as detecting differences between strains of a given CC, this technology facilitates recognition of putative areas of gene transfer (horizontal or vertical), resulting in variation amongst lineages. Furthermore, by studying more strains, recombination events may be more easily detected. Of course, this is dependant upon the design of the array and number of genes included in the compilation, as well as the strain collection under investigation. Table 4.1 displays the heatmap profile of the strain collection when grouped by clonal complex. By comparing the black areas, it can clearly be seen that there are many areas of similarity within a CC but which differ between CCs.

With respect to selecting markers for typing purposes, this may be easily achieved. However, rather than simply selecting genes present in a group, selecting sets of genes that are both present and absent provides a more discriminatory approach. Table 4.2 highlights some of those features most noticeable from the array data that may be used to define strains within a CC. However, with respect to the singletons (i.e. only 1 strain from a CC), more representative strains of these groups would need to be investigated on the array.

Table 4.1 Clonal Complexes

Snapshot profile of CGH results for a subset of the strains used in this study; strains were grouped into clonal complexes based on previous MLST data. Each horizontal strip represents one strain, whilst vertical strips represent a single gene profile; red areas mark gene presence, absence is indicated by green, and areas of ambiguity are marked in white (this is the same for all similar tables in this chapter to follow). The table shows that CGH similarities within a group are congruent with MLST data.

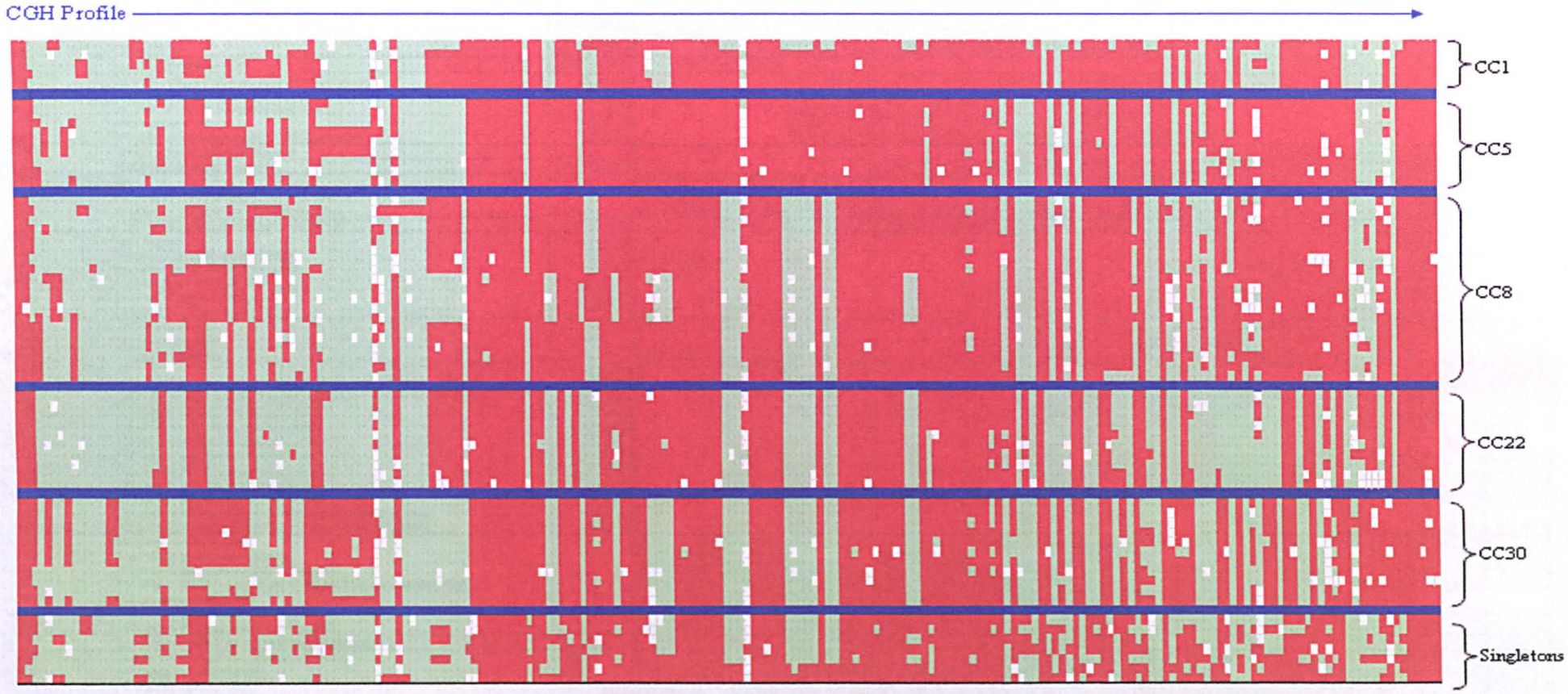


Table 4.2 Microarray features specific to the major clonal complexes analysed on the array. These represent the array features that may be used to define isolates within a CC. For example, SA2010-2012 are unique to the CC5 group, whilst SA0190-97 are found only in CC1 and CC5. Clonal complexes 1, 5 and 30 generally had one consensus profile for this genes list, whilst CC8 and CC22 had several variants. In particular, the ST239 and ST240 strains differed from the consensus CC8 profile at several loci.

		CC1	CC5	CC8 and variants							CC22 and variants					CC30
Gene ID	Gene Product	CC1	CC5	CC8-ST8	NCTC8325	F1 ST239-III	F4 ST239-III	F* ST239-III	F11 ST239	F9 ST240	CC22	F15-B3(0412)	F15-B*	F15 ST22	F15-B2*	CC30
E16-2025c SAR1903	V8PROTEASE	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
E162027c SAR1905	V8PROTEASE	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
E16-2979c SAR2789	SUBTILISIN (serine protease)	0	0	0	0	1	1	1	1	1	0	0	0	0	0	1
MW0062	HP similar to macrolide-efflux determinant [putative permease]	1	0	1	1	0	0	0	0	0	1	1	1	1	1	0
MW0063	portal protein	1	0	1	1	0	0	0	0	0	1	1	1			0
MW0064	HP, similar to transcriptional regulator (LysR family)	1	0	1	1	0	0	0	0	0	1	1	1	1		0
MW0065	terminase large subunit	1	0	1	1	0	0	0	0	0	1	1	1	1	1	0
MW0066	HP, similar to transcriptional regulator [LysR family protein]	1	0	1	1	0	0	0	0	0	1	1	1	1	1	0
MW0108	HP [cell wall surface anchor family protein]	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1
MW0120	truncated replication initiator protein	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
MW0171	HP	1	1	0	0	0	0		0	0	0	0	0	0	0	0
MW0172	HP, similar to ABC transporter ATP-binding protein	1	1	0	0	0	0	0	0	0	0	0		0	0	0
MW0178	HP, similar to ABC transporter ATP-binding protein	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
MW0263	conserved HP, similar to diarrheal toxin incomplete ORF	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
MW0552	major tail protein	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW0553	conserved HP [putative membrane protein]	1	0	0	1	1	1	1	1	1	0	0	0	0		0
MW0554	putative primase [putative membrane protein]	1	0	0	1	1	1	1	1	1	0	0	0	0		0
MW0558	conserved HP [putative membrane protein]	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW0559	conserved HP [putative membrane protein]	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW0560	HP [putative membrane protein]	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1047	HP, similar to exotoxin 1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1048	HP, similar to exotoxin 4	1	1	1	1	1	1	1	1	1	0	0	0		0	0
MW1049	HP, similar to exotoxin 3	1	1	0	1	1	1	1	1	1	0	0	0	1	1	0
MW1742	HP	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0
MW1744	HP	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0
MW1757	HP, similar to Ear protein (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1758:bsaG	HP, similar to epiG (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1759:bsaE	HP, similar to epiE (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1760:bsaF	HP, similar to epiF (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	1		0
MW1761	HP, similar to EpiP (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1762:bsaD	HP, similar to EpiD (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1763:bsaC	HP, similar to EpiC (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
MW1764:bsaB	HP, similar to EpiB (Genomic island νι Sa beta2)	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
MW1765	HP, similar to gallidermin precursor (bsaA1) (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1		0	0	1	0		0
MW1766:bsaA1	HP, similar to gallidermin precursor (bsaA1) (Genomic island νι Sa beta2)	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
SA0190	HP [putative membrane protein]	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA0191	conserved HP [putative membrane protein]	1	1	0	0		0		0		0	0	0	0	0	0
SA0192	HP, similar to ABC transporter ATP-binding protein	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA0193	HP, similar to Enterococcus faecalis plasmid pPD1 bacI [putative membrane protein]	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA0195	HP [putative membrane protein]	1	1	0	0	0	0	0	0	0	0	0	0			0
SA0196	conserved HP [putative membrane protein]	1	1	0	0	0	0		0	0	0	0	0	0	0	0
SA0197	HP, similar to ABC transporter ATP-binding protein	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA0276	conserved HP, similar to diarrheal toxin [diarrheal toxin (yukA)]	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
SA1633	probable beta-lactamase (Pathogenicity island SaPIn3)	0	1		0	0	0	0	0	0	0	1	0	0	0	0
SA1635	HP (Pathogenicity island SaPIn3)	0	1	0	0	0	0	0	0	0		0	0	0	0	0
SA1636	HP (Pathogenicity island SaPIn3)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA1640	conserved HP (Pathogenicity island SaPIn3)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA2010	HP, similar to RNA-directed DNA polymerase from retron EC86	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA2011	HP	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SA2012	HP	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

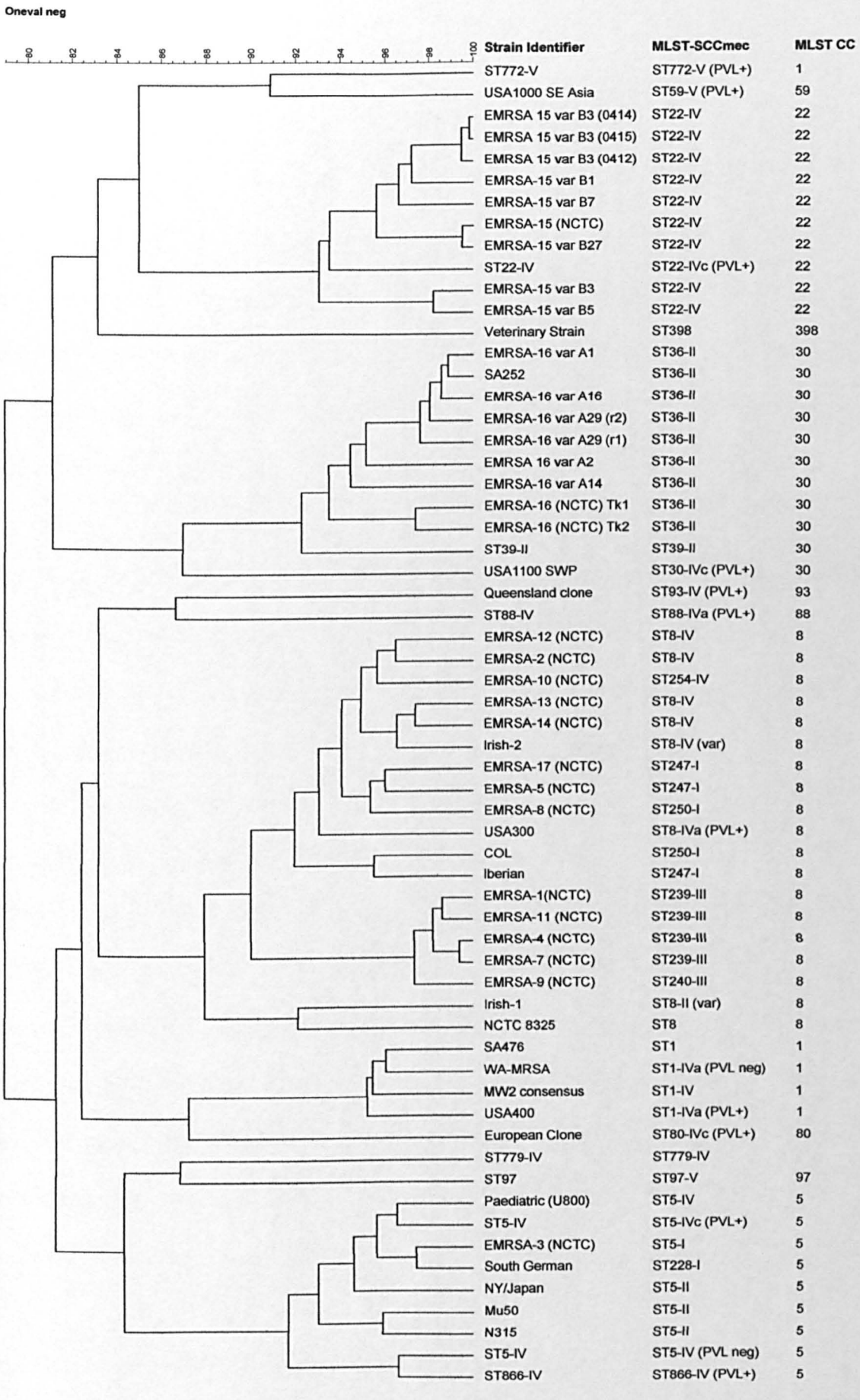


Figure 4.1 Clustering based on the complete array profiles for all the strains used in this study.

4.1.2 UK Epidemic HA-MRSA strains: EMRSA1-EMRSA17

4.1.2.1 Background

Epidemic MRSA strains (defined as "*those which have been identified in two or more patients in two or more hospitals*"; O'Neill *et al.*, 2001) have become a significant public health problem over recent years. One of the points of discussion and investigation into these strains is the reason behind their success within healthcare environments; they spread rapidly, and are difficult to control and eradicate.

EMRSA strains were originally recognised by phage typing characteristics (Marples *et al.*, 1986) and later confirmed as clonal by molecular typing techniques. The first of the epidemic MRSA strains recognised was dubbed EMRSA-1, and was detected in England in 1981. This clone became more widespread in the south east of England where it was the predominant clone until it began to decline in 1987 (Marples and Cooke 1988, Marples and Reith 1992). In this geographic location, EMRSA-1 was able to spread and persist in hospitals more successfully than other strain types. In the early 1990s, the national survey of MRSA showed EMRSA-1 was declining and the prevalence of EMRSA-3 was increasing (Marples and Reith, 1992). Cockfield and colleagues (2007) later noted that in any particular hospital, only one or two MRSA lineages predominate. Furthermore, international molecular epidemiological studies collected since the late 1980s suggest the massive geographic spread of MRSA is due to the dissemination of a relatively few epidemic clones (Crisostomo *et al.*, 2001).

In the late 1990s, EMRSA-15 (ST22-IV) and EMRSA-16 (ST36-II) became the predominant clones in UK hospitals and were estimated to account for greater than 95% of all MRSA bacteraemias in the UK (Johnson *et al.*, 2001). It seems reasonable to propose

that these clones possess genetic features enabling them to be particularly well adapted to survive and spread in hospitals compared with other MRSA. From a genomic point of view, the search for genes associated with this epidemic phenotype is a primary focus. Identification of such genes would be helpful in the identification of 'fit' strains and would increase understanding of the mechanisms of strain 'success' and pathogenicity. Almost 60% of the strains in this study were EMRSA strains of different genetic backgrounds; within this, almost half were representatives of EMRSA-15 and -16. The remainder included strains of EMRSA types 1 – 17 (except EMRSA-6 which produced poor hybridisation results). Analysis of these strains along with non-epidemic HA-MRSA sought to identify type-specific markers and to seek common features that might be responsible for, or at least correlate with, the EMRSA phenotype. Of the groups, several strains belong to the same CC and carry the same SCC*mec* type. The array profiles obtained from these strains are useful as a means of determining array typeability and discriminatory power. The typeability afforded by the array is determined by the ability to distinguish between similar isolates (many of which belong to the same CC). Furthermore, regarding the EMRSA-15 and EMRSA-16 PFGE subtypes, array discriminatory power is apparent by ability to separate different variants (subtypes) of these two clones (several subtypes of these groups were analysed). Finally, the data generated from this EMRSA strain collection should provide insights into traits possibly important in their success and transmissibility, with particular focus on the clones which have been predominant in the UK healthcare setting for the last 2 decades (EMRSA-15 and EMRSA-16).

4.1.2.2 Determination of array typeability and discriminatory power

The discriminatory ability of the array is illustrated in figure 4.2 and table 4.3 which show a tree diagram and heat map obtained from the profiles of EMRSA strains of CC8 (all strains excluding EMRSA-3, -15 and -16 which are not CC8). These show that the related

strains cluster well together, but can be distinguished from their near relatives. The specificity of the array is illustrated in table 4.4, which shows the ability to discriminate between the EMRSA-16 variants and other members of CC30.

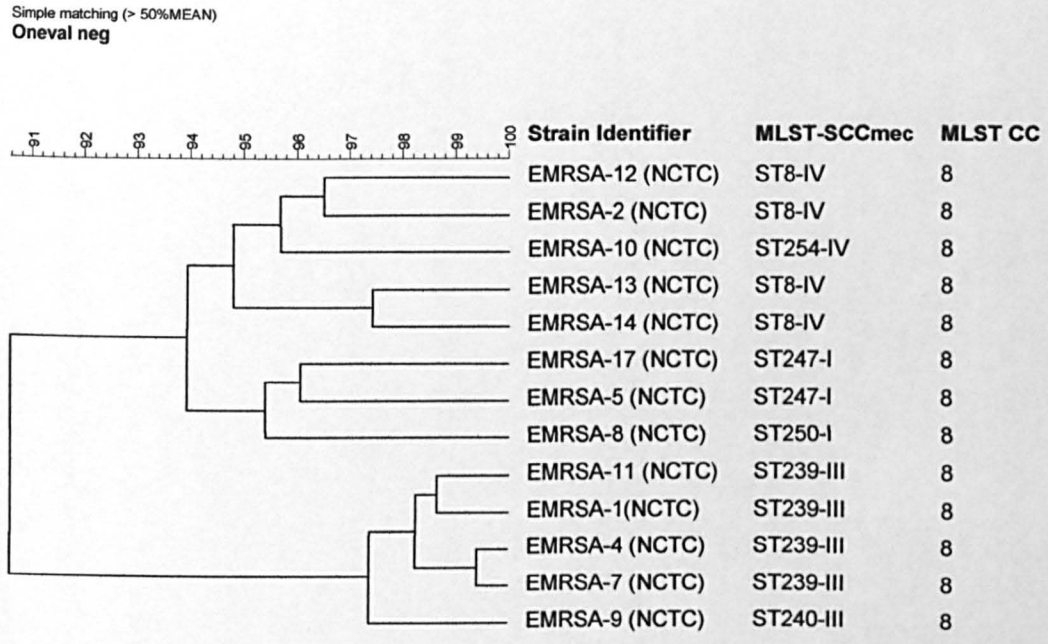


Figure 4.2 Dendogram of CC8 members of EMRSA1-EMRSA17. Clustering was based on Simple Matching and UPGMA.

Table 4.3 EMRSA1-EMRSA17 strains belonging to CC8

Partial profile of the CGH results of CC8 EMRSA strains. The strains of the same sequence type cluster together (compare vertical strips within a group and between groups).

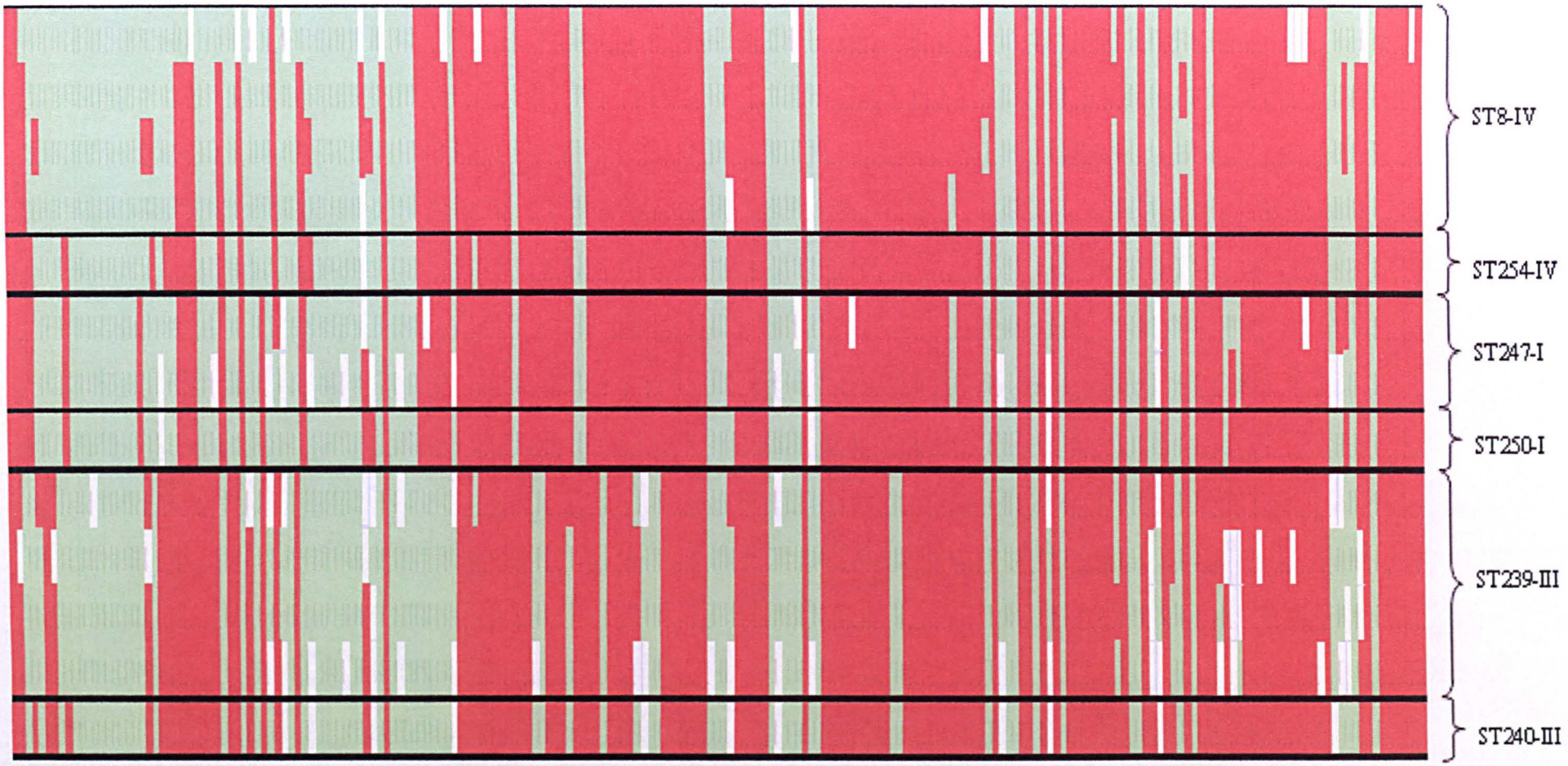
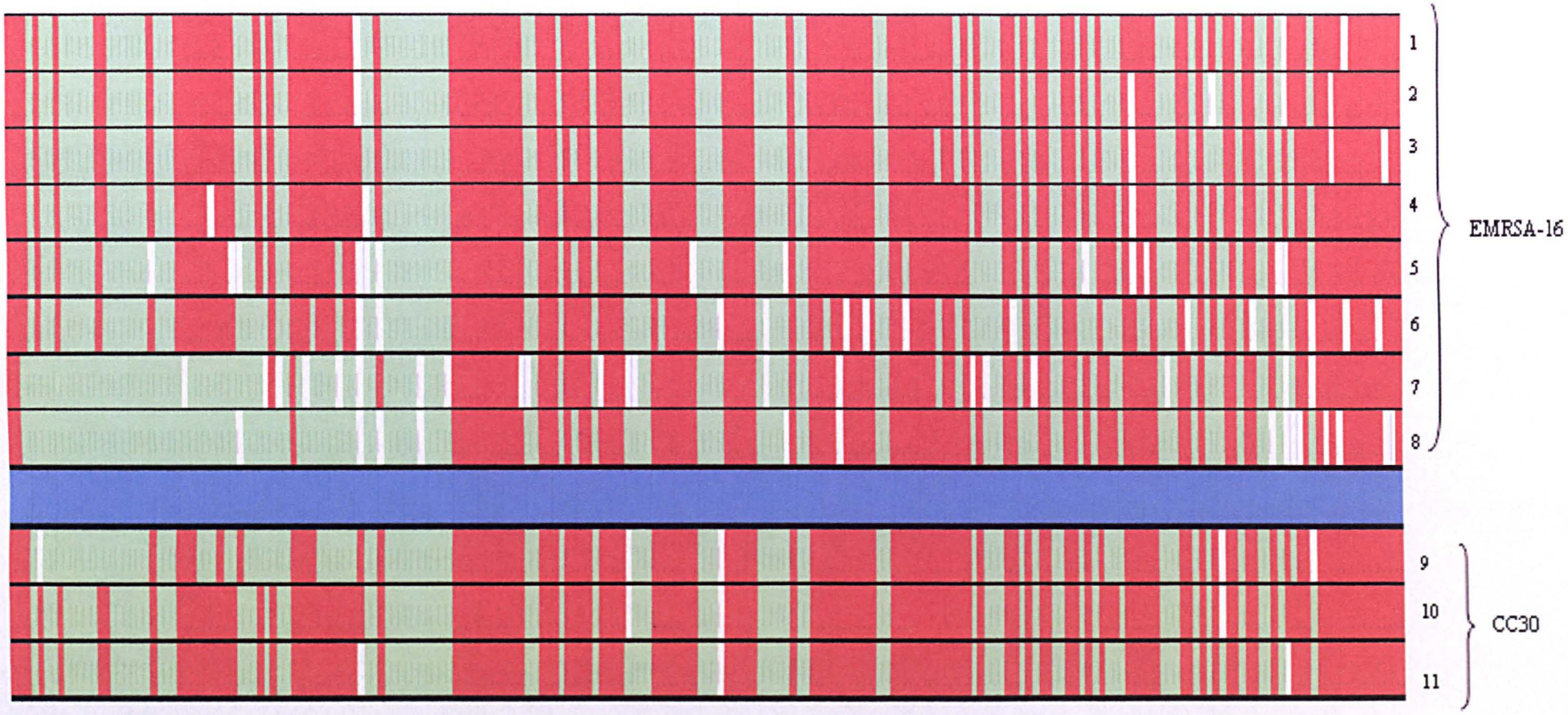


Table 4.4 CC30 strains

Comparison of EMRSA-16 variants with the other members of CC30. The top panel represents the EMRSA-16 classified strains; 1-6 correspond to differing pulsotypes of EMRSA-16 termed variants A1, A16, A29, A29 (replicate), A2 and A14 respectively, whilst 7 and 8 are technical replicates of the NCTC EMRSA-16 strain. The bottom panel denotes the other CC30 members in the following order (9, 10, and 11): USA1000 (ST30), ST39 and MRSA252 (ST36) respectively. The data show the strains display a general similarity, but variation is noted with the NCTC strains (7 and 8) and USA1000 (9). The complete sequence of MRSA252 (11; an EMRSA-16 representative) is known.



4.1.2.3 Investigating markers associated with success of HA strains

Moore and Lindsay (2002) studied EMRSA-15 and -16 strains by RFLP analysis (Southern blotting) and noted they were markedly different from each other, as well as from other EMRSA strains. They also noted EMRSA-3 was distinct from other UK EMRSA. Furthermore, EMRSA-1, -4, -7, -9 and -11 were classified as closely related. Similarly, EMRSA-2, -5, -6, -8, -10, -12, -13, -14 and -17 were also grouped. In addition, their analysis showed that EMRSA-15 and -16 and -3 all carried enterotoxin genes *seg* and *sei* which form part of the enterotoxin gene cluster (egc) that could play a role in virulence. Microarray-based analyses support these observations.

It can be said that it may be the absence rather than the presence of particular virulence genes that determines strain success or selective advantage. In such a case, the advantage may come from the reduced burden associated with the absence of these extra genes. The microarray data shows that EMRSA-15 and -16 strains differ from other EMRSA in the presence or absence of many genes (Tables 4.5 - 4.10). This accounts for the genetic distance between these strains and the remaining EMRSAs. The array study was designed to investigate whether there were common features of the EMRSA-15 and EMRSA-16 groups that account for their relative success. To address this, the EMRSA-15 and EMRSA-16 strains were compared to the EMRSA1-EMRSA17 group (Table 4.5). Features that differentiated the EMRSA-15/16 strains from the other groups, as well as from each other were determined (i.e. genes unique to EMRSA-15 and/–or -16). In the original array annotations, many of these genes were classified as 'hypothetical proteins'. Therefore, further *in silico* work was conducted to update these annotations by association to more recently sequenced *S. aureus* strains. The results are summarised in Tables 4.6-4.10.

Features unique to EMRSA-15 and EMRSA-16 strains (compared to the rest) were analysed. Table 4.6 & 4.7 show features that define EMRSA-15 strains compared with the remaining groups (including EMRSA-16). Few genes were found to be unique to the EMRSA-15s; a few lipoproteins and protease/peptidase genes were noted, as well as surface proteins. In comparison, the features that uniquely define the EMRSA-16 group from the rest were more extensive (Tables 4.8 & 4.9). Several protease genes, as well as enterotoxins, exotoxins, lipoproteins and genes associated with *S. aureus* pathogenicity island 1 (SaPln1) were present (Table 4.8). Furthermore, of those genes that were absent from the EMRSA-16 group only, several transcriptional regulators, ABC transporters, putative membrane proteins and toxins of the leukocidin family were identified. Of particular interest are the transcriptional regulatory genes that are different from those of the MW2 and N315 strains used in the oligo probe design. The implication here is one of a varied (or partly varied) regulatory network in the EMRSA-16 strains. Furthermore, also included are a block of genes encoding lantibiotic synthesis from genomic island nuSaα2.

Finally, features shared by EMRSA-15 and EMRSA-16 strains but absent from other strains were investigated (Table 4.10). The genes constituted purely the accessory genome. In particular, genes from bacteriophage ΦSa2mw (phage genes, but lacking MW1401 - large terminase), bacteriophage ΦN315 (enterotoxins), genomic island nuSaα (exotoxins), pathogenicity island SaPln3 (enterotoxins) and further enterotoxins from an unknown pathogenicity island. In view of the fact the core genomes being compared are distinct, EMRSA-16 strains were compared to the other CC30 strains, and EMRSA-15 to CC22 strains. These analyses showed ΦSa2mw genes to be unique to EMRSA-15 and EMRSA-16 strains (absent from the other members of CC22 and CC30 of which the CA-MRSA strains were included). However, the enterotoxins of SaPln3 were present in all the CC30 members; as was the case for the enterotoxins of ΦN315 in CC22 strains, i.e. these features were not unique to EMRSA-15. In conclusion, features unique to a particular

group of strains can be detected and selected from the array. Furthermore, it is apparent that the success of a particular clone is probably more to do with subtle differences between strains rather than any particular virulence factor since no one factor has been identified with certainty to date (although it is important to acknowledge such factors may not have been included in the array design). From the microarray data presented, it is tempting to speculate that one aspect of EMRSA15/16 success may be attributable to the acquisition of extra accessory genes, and in particular the ΦSa2mw genes.

Table 4.5 Comparison of EMRSA1-EMRSA17. Panel A shows all CC8 members. Panel B shows EMRSA-15 members (CC22) and panel C shows EMRSA-16 members (CC30). The blue panels separate the three groups – colours in the top blue panel highlight features specific to EMRSA-15 compared to the other two groups, whereas the lower blue panel highlight features specific to EMRSA-16 compared to the rest; a yellow panel highlights missing feature(s) whereas brown defines features unique to EMRSA-15 and/or EMRSA-16. The red areas indicate gene presence, green represent gene absence, and equivocal results remain white. The table provides an overview of the degree of similarity/differences between the groups; the details of this are summarised in Table 4.6 - 4.10.

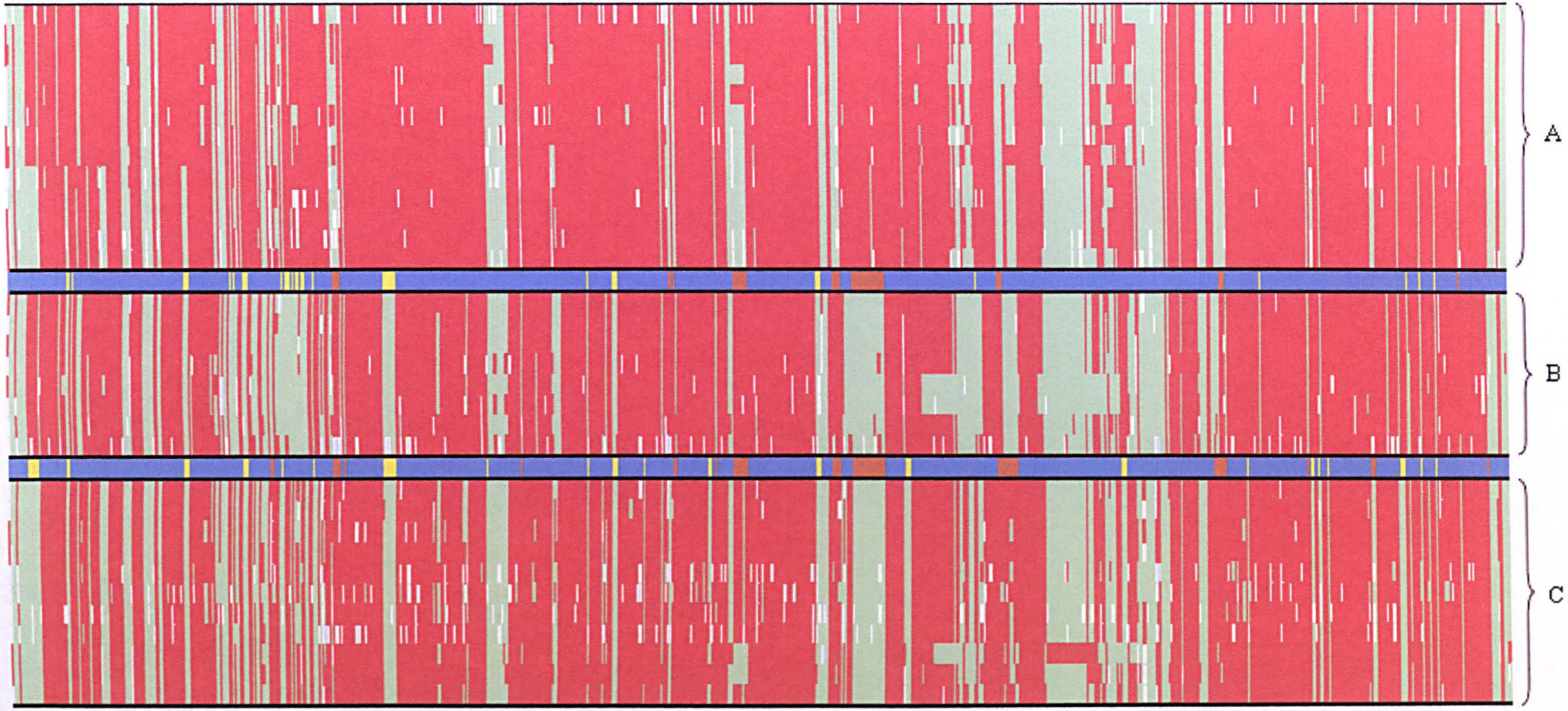


Table 4.6 CGH probes absent from EMRSA15 but present in EMRSA1-EMRSA16. Using a local BLAST utility, these genes were blasted against all available sequenced annotations for *S. aureus* strains. Homologies to genes of known function have been listed where possible (the same was conducted for Tables 4.7-4.10). Additional information obtained from the updated annotations are highlighted in grey.

Gene ID – Absent from EMRSA-15	Original annotation [Strain used for oligo design]	Homology to other genomes [strain]	E-value (%homology)
MW0108	hypothetical protein [MW2]	SAUSA300 0136 cell wall surface anchor family protein [USA300-FPR3757]	7.00E-22 (100)
		SAR0136 putative surface anchored protein (sasD) [MRSA252]	7.00E-22 (100)
SA0377	hypothetical protein [N315]	SAUSA300 0390 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
SA0378	hypothetical protein [N315]	SAV0417 hypothetical protein [Mu50]	7.00E-22 (100)
MW0400:lp13	hypothetical protein (lpl13) [MW2]	SAUSA300 0416 staphylococcus tandem lipoprotein [USA300-FPR3757]	7.00E-22 (100)
		SAR0444 putative lipoprotein [MRSA252]	7.00E-22 (100)
SA0400:lpl4	hypothetical protein (lpl4) [N315]	SAUSA300 0411 staphylococcus tandem lipoprotein [USA300-FPR3757]	7.00E-22 (100)
		SAS0400 putative membrane protein [MSSA476]	3.00E-12 (92)
		SACOL0481 staphylococcus tandem lipoprotein [COL]	3.00E-12 (92)
MW0401:lp14	hypothetical protein (lpl14) [MW2]	SAR0443 putative lipoprotein [MRSA252]	7.00E-22 (100)
		SAUSA300 0417 staphylococcus tandem lipoprotein [USA300-FPR3757]	3.00E-12 (97.37)
SA0404:lpl8	hypothetical protein (lpl8) [N315]	SAUSA300 0418 staphylococcus tandem lipoprotein [USA300-FPR3757]	4.00E-17 (96)
		SAR0444 putative lipoprotein [MRSA252]	4.00E-17 (96)
		SACOL0485 staphylococcus tandem lipoprotein [COL]	4.00E-17 (96)
SAV0416	hypothetical protein [Mu50]	SAUSA300 0390 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
SAV0417	hypothetical protein [Mu50]	SA0378 hypothetical protein [N315]	7.00E-22 (100)

SAV1811:splC	serine protease (splC) [Mu50]	SAUSA300 1756 serine protease SplC (splC) [USA300-FPR3757]	7.00E-22 (100)
		MW1753 serine protease SplC (splC) [MW2]	7.00E-22 (100)
SA2124:fosB	fosfomycin resistance protein fofB (fosB) [N315]	SAUSA300 2280 metallothiol transferase fosB (fosB) [USA300-FPR3757]	7.00E-22 (100)
		SAR2419 putative fosfomycin resistance protein (fosB) [MRSA252]	7.00E-22 (100)
SAV2451	subfamily S9C non-peptidase homologues [Mu50]	SAUSA300 2396 para-nitrobenzyl esterase (pnbA) [USA300-FPR3757]	7.00E-22 (100)
		SAR2541 putative carboxylesterase [MRSA252]	7.00E-22 (100)
		MW2375 hypothetical protein, similar to para-nitrobenzyl esterase chain A [MW2]	7.00E-22 (100)

Table 4.7 CGH probes present in EMRSA-15 but absent from EMRSA1-EMRSA16

Gene ID-Present in EMRSA-15	**Original annotation [Strain used for oligo design]**	**Homology to other genomes [strain]**	**E-value (%homology)**
SA1208	MW1263 hypothetical protein [MW2]	SA1208 hypothetical protein [N315]	7.00E-22 (100)
		SAS1316 putative membrane protein [MSSA476]	7.00E-22 (100)
SAV2596	family S1 unassigned peptidases	SA2382 truncated hypothetical protein, similar to glutamyl-endopeptidase [N315]	7.00E-22 (100)
		SAV2596 hypothetical protein [Mu50]	7.00E-22 (100)

Table 4. 8 CGH probes present in EMRSA-16 but absent from EMRSA1-EMRSA15, i.e. unique to EMRSA16

Gene ID – Present in EMRSA-16	Original annotation [Strain used for oligo design]	Homology to other genomes [strain]	E-value (%homology)
SA0396:lpl1	Hypothetical protein (lpl1) [N315]	SAUSA300 0410 staphylococcus tandem lipoprotein [USA300-FPR3757]	7.00E-22 (100)
		SAR0439 putative lipoprotein [MRSA252]	7.00E-22 (100)
MW0397:lp10	Hypothetical protein (lpl10) [MW2]	SAS0399 putative lipoprotein [MSSA476]	7.00E-22 (100)
		SAR0440 putative lipoprotein (pseudogene) [MRSA252]	3.00E-12 (92)
E16-0475	Exotoxin [MRSA252]	SAR0435 exotoxin [MRSA252]	7.00E-22 (100)
E16-0776	Putative membrane protein [MRSA252]	SAR0686 putative transposase (pseudogene) [MRSA252]	7.00E-22 (100)
E16-1213c	Exotoxin [MRSA252]	SAR1140 exotoxin [MRSA252]	7.00E-22 (100)
SA1819:tsst-1	Toxic shock syndrome toxin-1 (tst) [N315]	SAV2011 toxic shock syndrome toxin-1 (tst) [Mu50]	7.00E-22 (100)
		SAB0360c toxic shock syndrome toxin-1 (tsst-1) [RF122]	5.00E-20 (100)
E16-2042c	Enterotoxin [MRSA252]	SAB1701c enterotoxin O (seo) [RF122]	7.00E-22 (100)
		SAV1830 enterotoxin (seo) [Mu50]	4.00E-08 (89.36)
E16-2024c	Enterotoxin [MRSA252]	SAB1701c enterotoxin O (seo) [RF122]	7.00E-22 (100)
		SA1648 enterotoxin SeO (seo) [N315]	4.00E-08 (89.36)
		SAV1830 enterotoxin (seo) [Mu50]	4.00E-08 (89.36)
E16-2025c	Serine protease (pseudogene) [MRSA252]	SAR1903 serine protease (pseudogene) [MRSA252]	7.00E-22 (100)
E162027c	Serine protease [MRSA252]	SAR1905 serine protease [MRSA252]	7.00E-22 (100)
		SAB1670c serine proteinase (splE)[RF122]	3.00E-18 (100)
E16-2979c	Putative subtilase family protease [MRSA252]	SAR2789 putative subtilase family protease [MRSA252]	7.00E-22 (100)
SA1822	Hypothetical protein [N315]	SAB1896c hypothetical mobile element-associated protein [RF122]	7.00E-22 (100)
		SAB0357 bovine pathogenicity island protein Orf7 [RF122]	7.00E-22 (100)
		SACOL0904 pathogenicity island protein [COL]	7.00E-22 (100)
SA1826	Hypothetical protein [N315]	SAB1902c hypothetical mobile element-associated protein [RF122]	7.00E-22 (100)
		SAB0353 bovine pathogenicity island protein Orf11 [RF122]	7.00E-22 (100)
		SACOL0900 pathogenicity island protein [COL]	7.00E-22 (100)
SA1828	Hypothetical protein [N315]	SAR0375 hypothetical protein [MRSA252]	7.00E-22 (100)

		SAB1904c hypothetical mobile element-associated protein [RF122]	7.00E-22 (100)
SA1829	Hypothetical protein [N315]	SAV2022 hypothetical protein [Mu50]	7.00E-22 (100)
		SAB1905c hypothetical mobile element-associated protein [RF122]	4.00E-17 (96)
		SAB0350 bovine pathogenicity island protein Orf15 [RF122]	4.00E-17 (96)
SA1833	Hypothetical protein, similar to transcription regulator [Pathogenicity [N315]	SAV2026 hypothetical protein [Mu50]	7.00E-22 (100)
SA1834	Hypothetical protein [N315]	SAV2027 hypothetical protein [Mu50]	7.00E-22 (100)
SA1341	Hypothetical protein, similar to export protein SpcT [N315]	SAR1588 putative membrane protein [MRSA252]	2.00E-19 (98)
SA2259	Conserved hypothetical protein [N315]	SAV2471 conserved hypothetical protein [Mu50]	7.00E-22 (100)
SA2382	Truncated hypothetical protein, similar to glutamyl-endopeptidase [N315]	SAV2596 hypothetical protein [Mu50]	7.00E-22 (100)
SA2389	Truncated hypothetical protein, similar to metalloproteinase mpr [N315]	SAV2596 hypothetical protein [Mu50]	7.00E-22 (100)

Table 4.9 CGH probes absent from EMRSA-16 but present in other EMRSA1-EMRSA15

Gene ID - Absent EMRSA-16	Original annotation [Strain used for oligo design]	Homology to other genomes [strain]	E-value (%homology)
MW0062	Hypothetical protein, similar to macrolide-efflux determinant [MW2]	SAUSA300 0091 putative permease [USA300-FPR3757]	7.00E-22 (100)
MW0063	Conserved hypothetical protein [MW2]	SAUSA300 0092 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
MW0064	Hypothetical protein, similar to transcriptional regulator (LysR family) [MW2]	SAUSA300 0093 transcriptional regulator, LysR family domain protein [USA300-FPR3757]	7.00E-22 (100)
MW0065	Hypothetical protein [MW2]	SAUSA300 0094 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
MW0066	Hypothetical protein, similar to transcriptional regulator [MW2]	SAUSA300 0095 transcriptional regulator, LysR family domain protein [USA300-FPR3757]	7.00E-22 (100)
		SAS0066 LysR-family regulatory protein [MSSA476]	7.00E-22 (100)
SA0745	Hypothetical protein, similar to extracellular matrix and [N315]	SAV0814 hypothetical protein [Mu50]	7.00E-22 (100)
		SAB0747 truncated secreted von Willebrand factor-binding protein homolog [RF122]	2.00E-19 (98)
		MW0768 hypothetical protein, similar to extracellular matrix and plasma binding [MW2]	4.00E-17 (96)
		SAS0755 putative exported protein [MSSA476]	4.00E-17 (96)
SA1159	Hypothetical protein, similar to two-component response regulator [N315]	SAUSA300 1220 DNA-binding response regulator, LuxR family [USA300-FPR3757]	7.00E-22 (100)
		MW1209 hypothetical protein, similar to two-component response regulator [MW2]	7.00E-22 (100)
SA1320	Hypothetical protein [N315]	SAUSA300 1380 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
		SAB1349c probable lipoprotein [RF122]	4.00E-08 (89.36)
MW1742	Hypothetical protein [MW2]	SAS1724 hypothetical protein [MSSA476]	7.00E-22 (100)
MW1744	Hypothetical protein [MW2]	SAUSA300 1747 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)

		SAS1726 putative membrane protein [MSSA476]	7.00E-22 (100)
		SACOL1854 hypothetical protein [COL]	7.00E-22 (100)
MW2102	Hypothetical protein, similar to ferrichrome ABC transporter (permease) [MW2]	SAUSA300 2135 iron compound ABC transporter, permease protein [USA300-FPR3757]	7.00E-22 (100)
		SAB2057c ferrichrome ABC transporter [RF122]	7.00E-22 (100)
		SACOL2166 iron compound ABC transporter, permease protein [COL]	7.00E-22 (100)
SA2272	Hypothetical protein [N315]	SAV2484 hypothetical protein [Mu50]	7.00E-22 (100)
		SAUSA300 2427 conserved hypothetical protein, authentic frameshift [USA300-FPR3757]	3.00E-21 (100)
SA2286:sarH3	Staphylococcal accessory regulator A homolog (sarH3) [N315]	SAUSA300 2437 staphylococcal accessory regulator T (sarT) [USA300-FPR3757]	7.00E-22 (100)
		MW2417 staphylococcal accessory regulator A homolog (sarT) [MW2]	7.00E-22 (100)
SA2314	Hypothetical protein, similar to ABC transporter (ATP-binding [N315]	SAUSA300 2465 ABC transporter, ATP-binding protein [USA300-FPR3757]	7.00E-22 (100)
		SAB2400 probable ATP-binding ABC transporter [RF122]	7.00E-22 (100)
MW2446	Hypothetical protein, similar to ABC transporter (ATP-binding protein) [MW2]	SAUSA300 2465 ABC transporter, ATP-binding protein [USA300-FPR3757]	7.00E-22 (100)
		SAB2400 probable ATP-binding ABC transporter [RF122]	7.00E-22 (100)
MW2447	Conserved hypothetical protein [MW2]	SAUSA300 2466 putative membrane protein [USA300-FPR3757]	7.00E-22 (100)
MW1941	Hypothetical protein, similar to synergohymenotropic toxin precursor - *S.intermedius* [MW2]	SAUSA300 1974 Leukocidin/Haemolysin toxin family protein [USA300-FPR3757]	7.00E-22 (100)
MW1942	Hypothetical protein, similar to leukocidin chain lukM precursor [MW2]	SAUSA300 1975 Aerolysin/Leukocidin family protein [USA300-FPR3757]	7.00E-22 (100)
		SAB1876c probable leukocidin S subunit [RF122]	7.00E-22 (100)
		SA1813 hypothetical protein, similar to leukocidin chain lukM [N315]	7.00E-22 (100)

Table 4.10 Genes common to EMRSA-15 and EMRSA-16 but absent from other strains studied

Gene ID-Common to EMRSA15 & EMRSA16	Original annotation [Strain used for oligo design]	Homology to other genomes [strain]	E-value (%homology)
E16-1212c	SAR1139 exotoxin [MRSA252]	SAUSA300 1059 putative exotoxin 1 [USA300-FPR3757]	1.00E-04 (85.71)
E162040c	SAR1919 enterotoxin [MRSA252]	SAB1699c enterotoxin I (sei) [RF122]	7.00E-22 (100)
E16-2041c	SAR1920 enterotoxin [MRSA252]	SAB1700c sem-truncated (sem-truncated) [RF122]	7.00E-22 (100)
		SA1647 enterotoxin SEM (sem) [N315]	3.00E-09 (91.11)
E16-0463	SAR0424 exotoxin [MRSA252]	SAR0424 exotoxin [MRSA252]	7.00E-22 (100)
E16465	SAR0425 exotoxin [MRSA252]	SAR0425 exotoxin [MRSA252]	7.00E-22 (100)
E16-0467	SAR0427 exotoxin 3 (set3) [MRSA252]	SAR0427 exotoxin 3 (set3) [MRSA252]	7.00E-22 (100)
SA1643:sen	enterotoxin SeN (sen) [N315]	SAV1825 enterotoxin (sen) [Mu50]	7.00E-22 (100)
SA1644:yent2	enterotoxin YENT2 (yent2) [N315]	SAV1826 enterotoxin (yent2) [Mu50]	7.00E-22 (100)
		SAB1698c enterotoxin type C variant (sec-variant) [RF122]	1.00E-14 (95.65)
SA1646:sei	extracellular enterotoxin type I precursor (sei) [N315]	SAV1828 extracellular enterotoxin type I precursor (sei) [Mu50]	7.00E-22 (100)
		SAB1699c enterotoxin I (sei) [RF122]	2.00E-06 (93.75)
SA1647:sem	enterotoxin SEM (sem) [N315]	SAV1829 enterotoxin (sem) [Mu50]	7.00E-22 (100)
		SAB1700c sem-truncated (sem-truncated) [RF122]	1.00E-08 (89.58)
SAV1827:yent1	enterotoxin Yent1 (yent1) [N315]	SAV1827 enterotoxin (yent1) [Mu50]	7.00E-22 (100)
		SAB1698c enterotoxin type C variant (sec-variant) [RF122]	2.00E-19 (98)
SAV1830:seo	enterotoxin SeO (seo) [N315]	SAV1830 enterotoxin (seo) [Mu50]	7.00E-22 (100)
MW1767:LukD	leukotoxin (lukD) [MW2]	SAUSA300 1768 leukotoxin LukD (lukD) [USA300-FPR3757]	7.00E-22 (100)
		SAS1748 gamma-haemolysin component B precursor (hlgB) [MSSA476]	7.00E-22 (100)
MW1768:LukE	leukotoxin LukE (lukE) [MW2]	SAB1687c leukotoxin E subunit (lukE) [RF122]	7.00E-22 (100)
		SAS1749 gamma-haemolysin component A precursor (hlgA) [MSSA476]	7.00E-22 (100)
MW1754:splB	serine protease SplB (splB) [MW2]	SAUSA300 1757 serine protease SplB (splB) [USA300-FPR3757]	7.00E-22 (100)
MW1755:splA	serine protease SplA (splA) [MW2]	SAUSA300 1758 serine protease SplA (splA) [3.4.21.19] [USA300-FPR3757]	7.00E-22 (100)

MW1757	hypothetical protein, similar to Ear protein [MW2]	SAUSA300 1759 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)
		SAS1738 putative exported protein [MSSA476]	7.00E-22 (100)
MW1758:bsaG	hypothetical protein, similar to EpiG (bsaG) [MW2]	SAUSA300 1760 lantibiotic epidermin immunity protein F (epiG) [USA300-FPR3757]	7.00E-22 (100)
		SAS1739 putative lantibiotic ABC transporter protein [MSSA476]	7.00E-22 (100)
MW1759:bsaE	hypothetical protein, similar to EpiE (bsaE) [MW2]	SAUSA300 1761 lantibiotic epidermin immunity protein F (epiE) [USA300-FPR3757]	7.00E-22 (100)
		SAS1740 putative lantibiotic ABC transporter protein [MSSA476]	7.00E-22 (100)
		SAOUHSC 01947 membrane protein, putative [NCTC 8325]	7.00E-22 (100)
MW1760:bsaF	hypothetical protein, similar to EpiF (bsaF) [MW2]	SAUSA300 1762 lantibiotic epidermin immunity protein F (epiF) [USA300-FPR3757]	7.00E-22 (100)
		SAS1741 putative lantibiotic transport ATP-binding protein [MSSA476]	7.00E-22 (100)
MW1761	hypothetical protein, similar to EpiP precursor (bsaP) [MW2]	SAUSA300 1763 lantibiotic epidermin leader peptide processing serine protease EpiP (epiP)[USA300-FPR3757]	7.00E-22 (100)
		SAB1679c probable serine protease precursor [RF122]	7.00E-22 (100)
		SAS1742 putative lantibiotic leader peptide processing serine protease [MSSA476]	7.00E-22 (100)
MW1762:bsaD	hypothetical protein, similar to EpiD (bsaD) [MW2]	SAUSA300 1764 lantibiotic epidermin biosynthesis protein EpiD (epiD) [4.1.1.36] [USA300-FPR3757]	7.00E-22 (100)
		SAS1743 putative lantibiotic modifying enzyme [MSSA476]	7.00E-22 (100)
		SACOL1875 epidermin biosynthesis protein EpiD (epiD) [COL]	7.00E-22 (100)
MW1763:bsaC	hypothetical protein, similar to EpiC (bsaC) [MW2]	SAUSA300 1765 lantibiotic epidermin biosynthesis protein EpiC (epiC) [USA300-FPR3757]	7.00E-22 (100)
		SAS1744 putative lantibiotic biosynthesis protein [MSSA476]	7.00E-22 (100)
		SACOL1876 epidermin biosynthesis protein EpiC, authentic point mutation (epiC) [COL]	7.00E-22 (100)
MW1764:bsaB	hypothetical protein, similar to EpiB (bsaB) [MW2]	SAUSA300 1766 lantibiotic epidermin biosynthesis protein EpiB (epiB) [USA300-FPR3757]	7.00E-22 (100)
		SAS1745 putative lantibiotic biosynthesis protein [MSSA476]	7.00E-22 (100)
		SACOL1877 epidermin biosynthesis protein EpiB (epiB) [COL]	7.00E-22 (100)

MW1765	hypothetical protein, similar to gallidermin precursor (bsaA2) [MW2]	SAUSA300 1767 lantibiotic epidermin biosynthesis protein EpiA (epiA) [USA300-FPR3757]	7.00E-22 (100)
		SACOL1878 lantibiotic epidermin precursor EpiA (epiA) [COL]	7.00E-22 (100)
		SAOUHSC 01953 Gallidermin superfamily epiA, putative [NCTC 8325]	7.00E-22 (100)
MW1766:bsaA1	hypothetical protein, similar to gallidermin precursor (bsaA1) [MW2]	SAS1747a lantibiotic precursor [MSSA476]	7.00E-22 (100)
		SAB1685c hypothetical protein [RF122]	3.00E-18 (97.92)
MW1386	hypothetical protein [MW2]	SAUSA300 1389 phiSLT ORF636-like protein [USA300-FPR3757]	7.00E-22 (100)
		SAR1503 hypothetical phage protein [MRSA252]	7.00E-22 (100)
		SAOUHSC 01521 SLT orf 636-like protein [NCTC 8325]	7.00E-22 (100)
MW1390	hypothetical protein [MW2]	SAUSA300 1393 phiSLT ORF2067-like protein, phage tail tape measure protein [USA300-FPR3757]	7.00E-22 (100)
		SACOL0379 prophage L54a, tail tape meausure protein, TP901 family [COL]	7.00E-22 (100)
		SAOUHSC 01525 phage tail tape meausure protein, TP901 family, core region domain protein [NCTC 8325]	7.00E-22 (100)
MW1393	MW1393 major tail protein [MW2]	SAUSA300 1397 phiSLT ORF213-like protein, major tail protein [USA300-FPR3757]	2.00E-21 (100)
		SAS0940 major tail protein [MSSA476]	2.00E-21 (100)
		SACOL0375 prophage L54a, major tail protein, putative [COL]	2.00E-21 (100)
		SAOUHSC 01529 major tail protein [NCTC 8325]	2.00E-21 (100)
MW1400	hypothetical protein	SAUSA300 1403 phiSLT ORF412-like protein, portal protein [USA300-FPR3757]	2.00E-21 (100)
	MW1400 portal protein [MW2]	SAR1518 portal protein [MRSA252]	2.00E-21 (100)
MW1401	MW1401 terminase large subunit [MW2]	SAOUHSC 01538 phage terminase, large subunit, putative [NCTC 8325]	7.00E-22 (100)
		SAUSA300 1404 phiSLT ORF 563-like protein, terminase, large subunit [USA300-FPR3757]	4.00E-08 (89.36)
		SACOL0367 prophage L54a, terminase, large subunit, putative [COL]	4.00E-08 (89.36)
MW1405	MW1405 hypothetical protein [MW2]	SAUSA300 1408 phage helicase [USA300-FPR3757]	7.00E-22 (100)
MW1408	MW1408 hypothetical protein [MW2]	SAUSA300 1410 virulence-associated protein E [USA300-FPR3757]	7.00E-22 (100)
MW1769	MW1769 hypothetical protein [MW2]	SAUSA300 1770 conserved hypothetical protein [USA300-FPR3757]	7.00E-22 (100)

4.2 Investigating recombination events

4.2.1 Background

Genetic exchange is an important part of the evolution of any organism, facilitated by transformation, transduction and/or conjugation. Transduction, the most common of these mechanisms, is mediated by bacteriophage, and can result in the replacement of tens of kilobases *in vitro* (Milkman *et al.*, 1999). Conjugation, also mediated by mobile genetic elements, involves cell-to-cell contact. Conjugative plasmids or transposons facilitate the movement of host DNA, which can be hundreds of kilobases *in vitro* (Lloyd and Buckman 1995, Milkman *et. al.* 1999). Transformation on the other hand, involves the uptake of DNA from the local environment, by competent bacteria, but results in smaller replacements of less than ten kilobases in nature (Feil *et. al.*, 2000). These mechanisms of horizontal gene transfer have played an important role in the genetic diversity and epidemiology of pathogenic bacteria. Although not all such genetic events result in enhanced fitness, recognition of these events is of great interest (genetic exchange resulting in loss of fitness is unlikely to be detected). In the *S. aureus* genome, the accumulation of virulence and antibiotic resistance genes has been mediated via bacteriophage, transposons and plasmids alike. However, these genetic changes have not been limited to the accessory genome. The microarray was used to probe for recombination events that may have occurred during the evolution of successful *S. aureus* lineages.

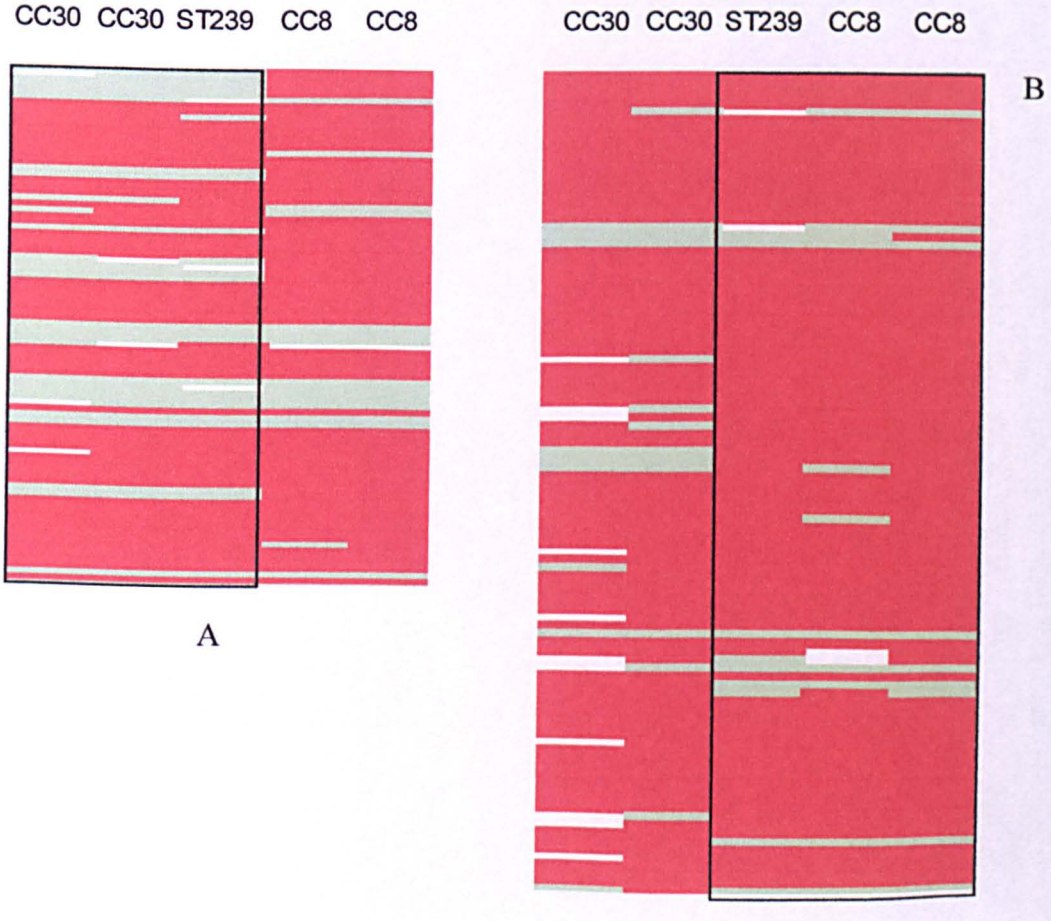
4.2.2 ST239 lineage

It has been reported (Robinson and Enright, 2004) that ST239 (e.g. EMRSA-1) is a recombinant lineage. The evidence for this was derived from MLST-based analysis, but it was also revealed using the original version of the virulence-associated microarray

(Saunders *et al.* 2004). ST239 appears to have originated from a large-scale recombination event between a CC30 strain (which contributed approximately 0.3Mb, including the origin of replication) and a CC8 strain (which contributed the remaining genes, approx. 2.5Mb). This is illustrated in Table 4.11 which shows heat maps for ST239, two CC30 strains and two CC8 strains. Table 4.11a shows genes around the origin of replication and similarity between ST239 and CC30. Table 4.11b shows genes remote from the origin and congruence between ST239 and CC8. Strain ST239 has been a significant public health problem in terms of its success and transmission, causing many hospital infections world-wide.

Table 4.11 Similarities between ST239 and CCs 8 and 30 indicate recombination.

Genes present are shown in black, genes absent in grey and no result in white. Each vertical strip represents results for a subset of probes (genes) in a single strain. The left (A) and right (B) panels represent genes taken from near to the origin of replication and remote from the origin of replication respectively. For simplicity results for only five isolates are shown, two different CC30 strains, two CC8s and one ST239 (CC8).



4.2.3 ST22 lineage

In the same way that the origin of ST239 may be illustrated using heat maps (Table 4.11) similarities were identified between the core genome of strains of CC22 and representatives of other clonal complexes (data not shown). These relationships are represented in the dendrograms shown in figure 4.3. These indicate that CC22, which includes the EMRSA-15 lineage prevalent in the UK, evolved via a large scale recombination between CC8-like and CC30-like strains. Approximately 0.9MB from around the origin of replication (between genomic islands nuSa4 and nuSaα, figure 4.4) appears to have originated from CC8 with the remainder from CC30. This is illustrated in figure 4.3 which shows dendrograms representing similarities between CC8, CC22 and CC30 strains based on the two sets of contiguous core genes. Figure 4.3a shows that CC8 and CC22 cluster based on genes around the origin of replication (nuSa4 to nuSaα) while figure 4.3b shows CC22 and CC30 clustering based on the remaining genes.

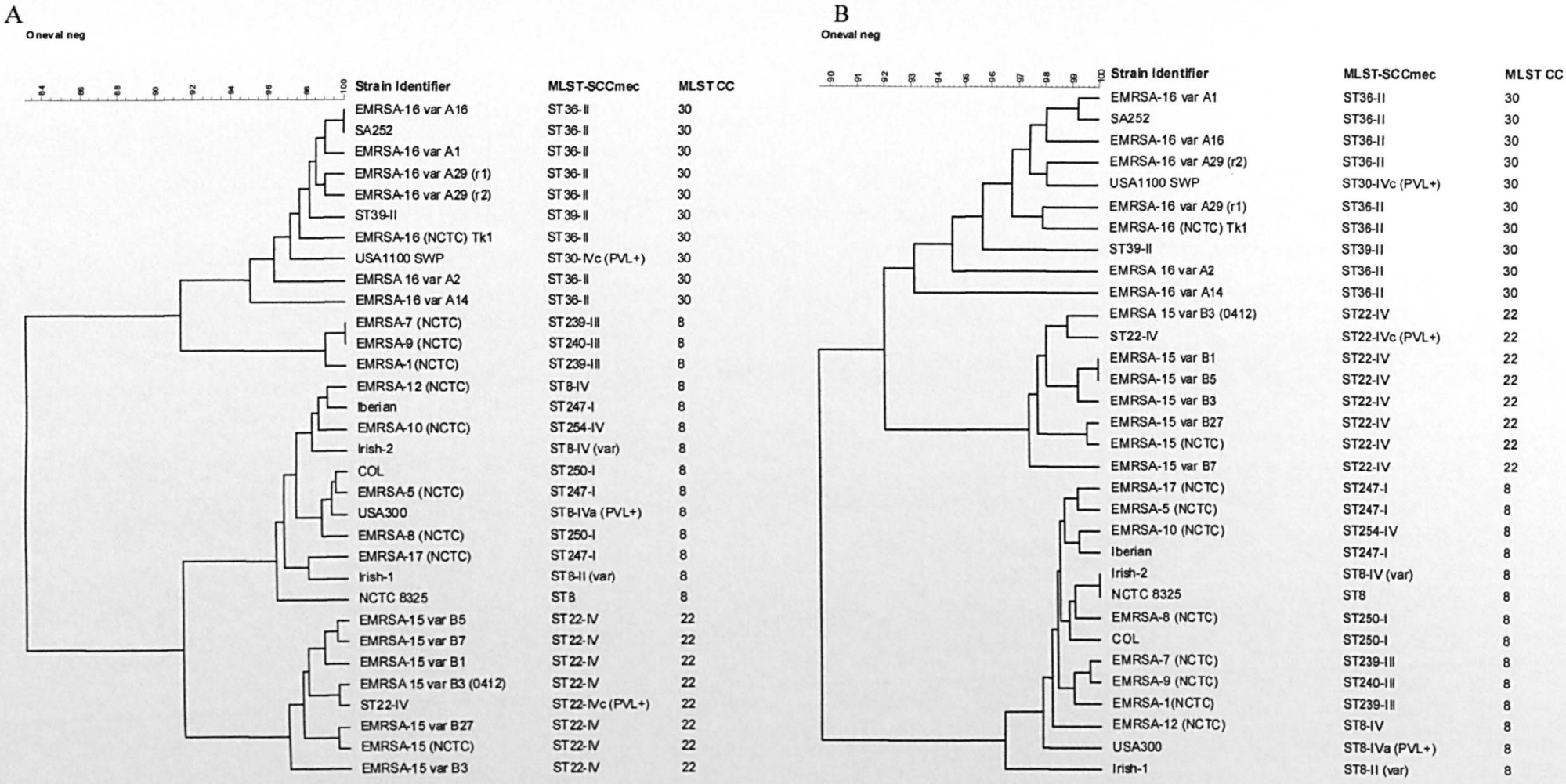


Figure 4.3 Dendograms indicating that CC22 may have been derived by large-scale recombination. The dendrograms represent similarities derived only from core genes for all strains belonging to CCs 8, 22 and 30. Panel A shows the result obtained for genes situated clockwise between genomic islands nuSa4 and nuSaα (including the origin of replication); here CC22 clusters with CC8 strains. Panel B shows the result for the remaining genes; CC22 clusters with CC30 group.

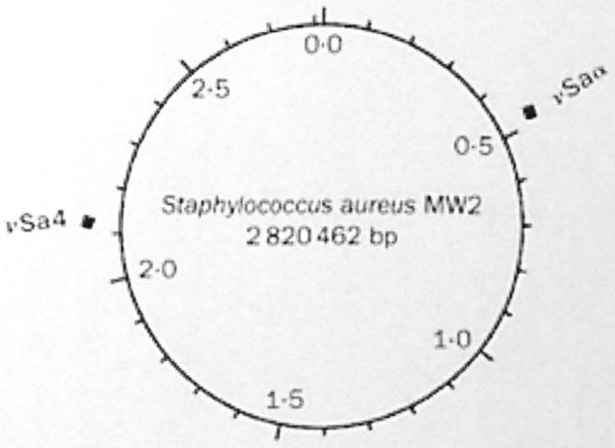


Figure 4.4 Chromosomal map of the MW2 genome used to show the location of the genomic islands nuSa4 and nuSaα. ST22 is proposed (in this thesis) to be a recombinant strain derived from a large scale recombination event between a CC8 and a CC30 background. The area around the origin of replication between nuSa4 and nuSaα is proposed to have originated from the CC8 background, with the remainder from CC30.

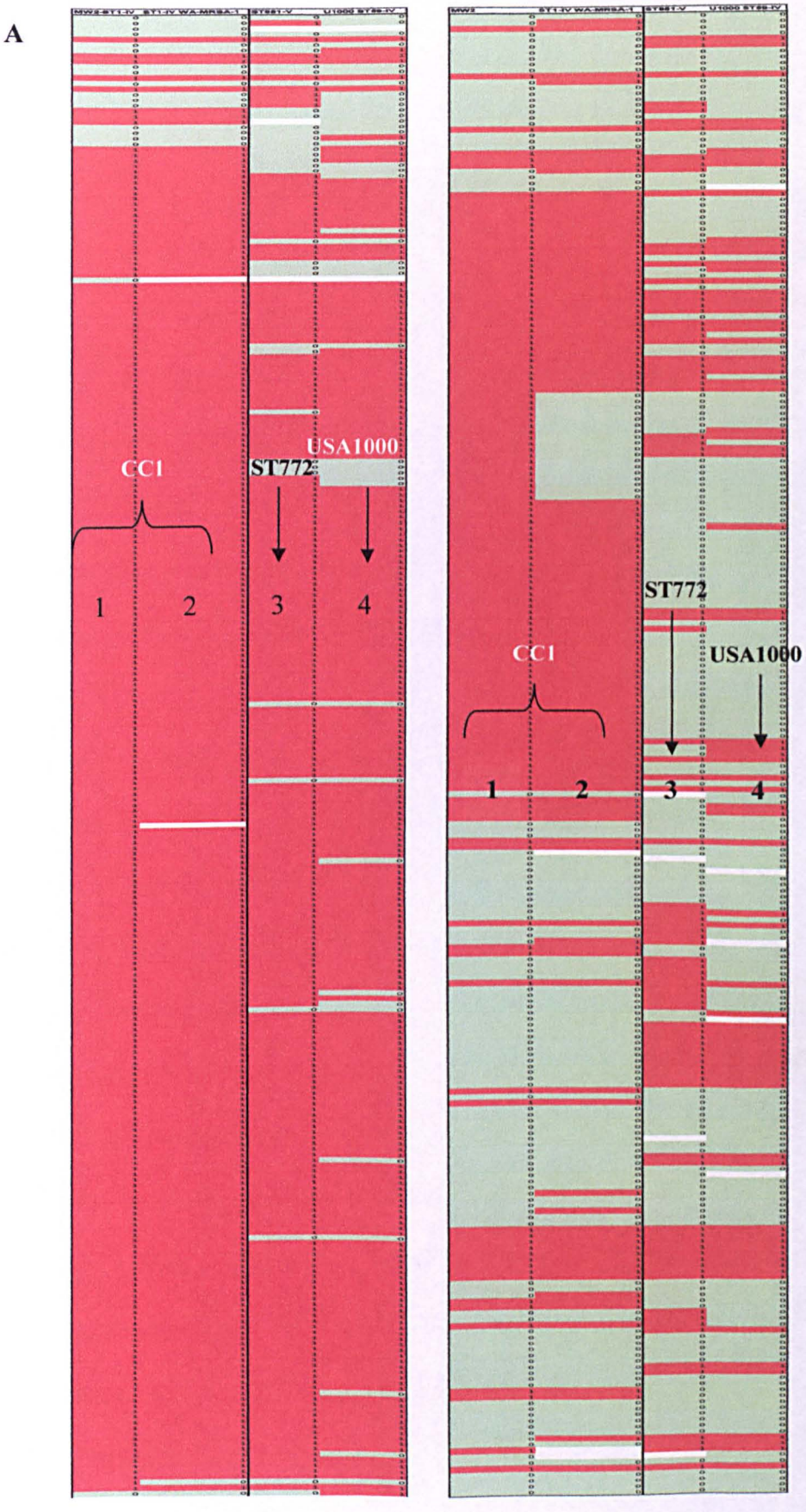
4.2.4 ST772 lineage: a unique example of clonal diversification

A further example of genetic exchange can be postulated for the ST772 lineage (represented by the Bengal-Bay clone in this data set; Ellington *et al.*, 2009), which belongs to CC1. Its MLST allelic profile shows it to be a single locus variant from the ST1 strains of CC1 (Table 4.12). From this point of view, these strains should all have a very similar genetic background, which in the case of the CC1 strains, the array data demonstrates this similarity (Table 4.13). However, the CGH data show the ST772 clone presents a different profile to the CC1 strains. In fact, ST772 appears to be more similar to USA1000 (CC59, which has no MLST alleles in common to ST772, Table 4.12) than to members of CC1. This overall genetic similarity can be seen in figure 4.1 (tree diagram based on the complete genetic profile). Based on prior information, the only commonality between these isolates was carriage of the type V SCC*mec* element, and geographic association with the Asian sub-continent (ST772 Bangladesh, and ST59 S.E. Asia, Taiwan). This was investigated further using the heatmaps. Table 4.13 shows ST772 compared with USA1000 and members of CC1 (MW2 and WA-MRSA) over both the core and accessory genomes. In both cases the greater similarity of ST772 to USA1000 can be clearly seen.

Table 4.12 MLST allelic profiles for *S.* aureus sequence types ST1, 772 and 59. ST772 is a single locus variant to ST1 (at the *pta* locus). ST59 shares no alleles in common to ST772.

MLST type	arcC	aroE	glpF	gmk	pta	tpi	yqiL
ST1	1	1	1	1	1	1	1
ST772	1	1	1	1	22	1	1
ST59	19	23	15	2	19	20	15

Table 4.13 Similarities between ST772 [3] and ST59 (USA1000) [4] suggest recombination. ST772 and USA1000 were compared to members of CC1 (MW2 [1] and WA-MRSA [2]) over core (panel A) and accessory genes (panel B). ST772 is a member of CC1, although the array data show it to cluster with USA1000 (ST59) based on the core and accessory genome.



From prior knowledge of S. *aureus* evolution, the possibility a lineage with the same MLST alleles as a particular clonotype (ST772 and CC1) would present such large genetic differences over the core and accessory genome, seems unlikely. Had the variation been limited to just the accessory genome, this would have been accounted simply by the gain/loss of mobile genetic elements. Furthermore, had variation been seen in several MLST alleles (instead of 1 allele alone), a recombination event could have been a possibility. However, ST772 appears to be a genetically unusual. The array shows that its core genome as well as the accessory genome is unlike ST1. Additionally, it is not an obvious mosaic like the ST239 or ST22 lineages. From the perspective of molecular typing, this shows that MLST can be misleading in inferring degree of genetic similarity. That is to say, this typing method (which has been the framework for studying *S. aureus*) is not always a reliable indication of genotype, and this can only be highlighted by looking at the overall genomic profile rather than a few select genes.

A possible theory for the relationship between ST772 and ST1 is descent from a common progenitor. Additionally, ST59 could have descended from ST772; these relationships are summarised in the diagram below (fig. 4.5). Here it is proposed that multiple recombination events (involving core non-essential genes and accessory genes) in the ST1 MSSA background (over a short period of time) gave rise to the ST772 group without affecting the core-essential genes. Following on from this, the accumulation of point mutations in the core-essential genes (including the MLST alleles) in the ST772 background gave rise to ST59. It is speculated that the acquisition of type V SCC*mec* could have occurred either after the point mutations or during the period of recombinational exchanges. Indeed this explanation would explain why ST59 and ST772 are more similar than to ST1, in that they are more genetically linked in their evolutionary descent, compared with ST1.

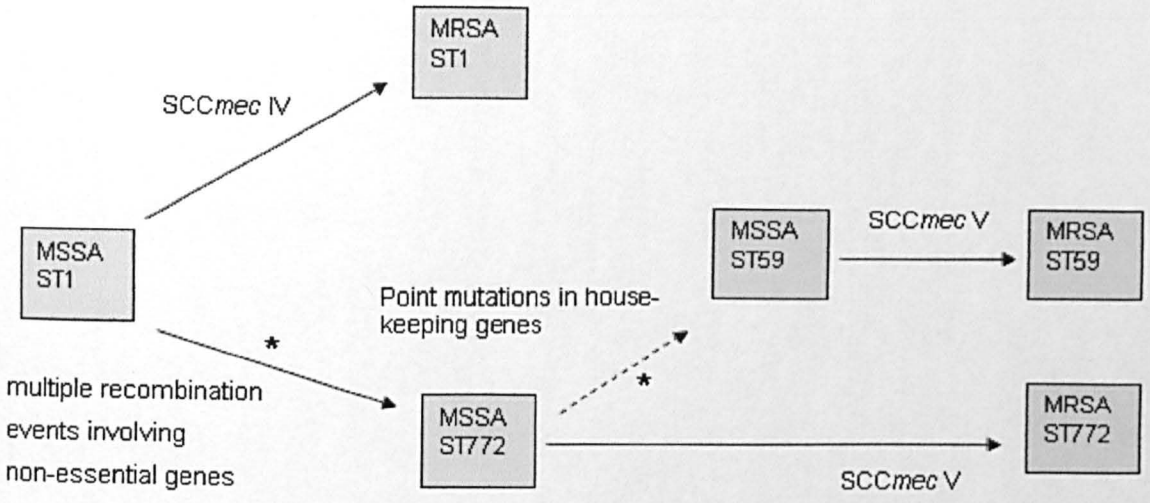


Figure 4.5 Proposed relationship between ST1, ST772 and ST59.
ST772 is proposed to have descended from the ST1 genetic background which would explain the commonality of 6 out of 7 MLST alleles between these two clones. The accumulation of multiple recombination events over the non-essential core genes (as well as the accessory genome) over a short period of time, followed by point mutations in the core-essential genes could explain the linkage between ST1, ST772 and ST59. ' *' marks alternative points at which SCC*mec* type V may have been acquired.

These data show that the analysis of many genes of an organism has more benefit in providing a more complete picture of the genomic content of an organism and thus for evolutionary studies. In some instances, the MLST framework for analysing evolutionary relationships can be misleading. Furthermore, the evolutionary picture of successful *S. aureus* strains (such as ST239 and ST22) may be more complex than originally anticipated. Large scale chromosomal exchange, as well as the accumulation of virulence factors could be the key to the spread and success of these clones.

4.3 Exploring genotypic differences that may explain variation in epidemiology and pathogenicity of *S. aureus*

4.3.1 Distribution of genes between HA and CA *S. aureus*

One of the aims of this study was to identify any patterns of genotypic variation that might account for differences between the HA and CA phenotypes. Initially, genotypically defined groups of HA and CA-strains, belonging to the same clonal cluster and harbouring the same SCC*mec* type were analysed and compared. This strain collection included three clusters (Tables 4.14 - 4.16) belonging to CCs 5, 8 and 22. The results showed that the core genome was well conserved even between the groups. However, as expected, variability was seen with respect to the mobile genetic elements, and in particular the bacteriophage genes. Bacteriophage are recognised as key factors in bacterial pathogenesis (Boyd *et al.* 2001, Wagner and Waldor 2002) and bacterial population dynamics (Burroughs *et al.* 2000, Hendrix 2003). Given that the core genome of *S. aureus* strains are relatively similar across clonal complexes, variation in the genomic content of the mobile genetic elements (see Appendix I) are likely to be important influences on the pathogenicity potential and success of particular strains.

For the CC5 strains (Table 4.14), none of the CA strains carried the genes for ΦSa2mw (MW1390-MW1405), whereas two of these three strains carried a group of consecutive genes for ΦSa3mw (MW1896-MW1912) which were absent in the HA-strain. Additionally, these same two CA-strains carried a set of genes from SaPI1 (*S. aureus* pathogenicity island 1), which were also absent in the HA strain.

The HA strains of CC8 were represented by four strains, Irish-2 and three EMRSA strains (EMRSA-2, -13 and -14). This group of strains were very similar with respect to the core genome and most parts of the accessory genome (some differences were noted on Tn554, ΦSa2mw and nuSaα). Table 4.15 compares these strains with USA300 (CA-strain) and NCTC 8325 (MSSA strain). Here USA300 mainly differed with respect to the *arcBDAR* genes (USA62/64-66 respectively, located within the SCC*mec* region), whereas the MSSA strain displayed a different profile over the ΦSa3mw and ΦN315 loci.

In the comparison of the CC22 strains, the profiles for the CA-strain and the HA-strains (three HA-strains taken from the same clinical patient) were clearly distinct over the loci of ΦSa2mw and ΦSa3mw (Table 4.16).

Table 4.14 CC5 HA- versus CA-strains.

Illustrated are the main features on the array that showed block variation between the profiles of the HA and CA strains of CC5. These were in the mobile genetic elements ΦSa2mw, ΦSa3mw and SaPln1. Other subtle differences were observed over the remaining profile (data not shown).

Loci	Gene ID	Product Description	Paed U800 ST5-IV	ST5-IV B3	ST5-IV PVL	ST866-IV
	MW1390	subfamily M23B unassigned peptidases (MW1390 protein)	1	0	0	0
Phage phi Sa 2mw	MW1393	conserved hypothetical protein	1	0	0	0
Phage phi Sa 2mw	MW1400	hypothetical protein	1	0	0	0
Phage phi Sa 2mw	MW1401	hypothetical protein	1	0	0	0
Phage phi Sa 2mw	MW1405	hypothetical protein	1	0	0	0
Phage phi Sa 3mw	MW1896	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1897	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1898	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1899	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1900	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1901	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1902	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1903	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1904	hypothetical protein [Capsid protein]	0	1	0	1
Phage phi Sa 3mw	MW1905	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1906	portal protein (phage portal protein)	0	1	0	1
Phage phi Sa 3mw	MW1907	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1908	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1909	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1910	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1911	hypothetical protein	0	1	0	1
Phage phi Sa 3mw	MW1912	hypothetical protein	0	1	0	1
PIsland SaPln1	SA1819:tsst-1	toxic shock syndrome toxin-1	0	1	0	1
PIsland SaPln1	SA1822	hypothetical protein	0	1	0	1
PIsland SaPln2	SA1826	hypothetical protein	0	1	0	1

HA CA

Table 4.15 CC8 HA- versus CA-MRSA and MSSA-strains. Illustrated are the main features on the array that showed similarities and/or differences between the profiles of MSSA and MRSA (HA and CA) strains of CC8. These were again localised to the accessory genome.

Loci	Gene ID	Product Description	EMRSA-2	EMRSA-13	EMRSA-14	IrirL-2	USA300	NCTC 8325
Girland nu Sa alpha3mu	MW0751	hypothetical protein	1	0	0	0	1	0
Girland nu Sa alpha3mu	MW0754	hypothetical protein	1	0		0	1	0
Girland nu Sa alpha3mu	MW0755	hypothetical protein	1	0	0	0	1	0
Phage phi Sa 2mu	MW1386	conserved hypothetical protein	1	0	1	0	1	1
Phage phi Sa 2mu	MW1390	subfamily M23B unassigned peptidases (MW1390 protein)	1	0	1	0	1	1
Phage phi Sa 2mu	MW1393	conserved hypothetical protein	1	0	1	0	1	1
Phage phi Sa 2mu	MW1400	hypothetical protein	1	0	1	0	1	1
Phage phi Sa 2mu	MW1401	hypothetical protein	1	0		0	0	1
Phage phi Sa 2mu	MW1405	hypothetical protein	1	0	1	0	1	1
Phage phi Sa 2mu	MW1408	hypothetical protein	1	0	1	0	1	1
Phage phi Sa 3mu	MW1895	subfamily M23B unassigned peptidases (MW1895 protein)	0	0	0	0	0	1
Phage phi Sa 3mu	MW1896	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1897	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1898	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1899	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1900	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1901	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1902	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1903	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1904	hypothetical protein [Capsid protein]	0	0	0	0	0	1
Phage phi Sa 3mu	MW1905	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1906	portal protein (phage portal protein)	0	0	0	0	0	1
Phage phi Sa 3mu	MW1907	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1908	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1909	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1910	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1911	hypothetical protein	0	0	0	0	0	1
Phage phi Sa 3mu	MW1912	hypothetical protein	0	0	0	0	0	1
Phage phiN315	SA1766	hypothetical protein	1	1	1	1	1	0
Phage phiN315	SA1768	hypothetical protein	1	1	1	1	1	0
Phage phiN316	SA1769	hypothetical protein	1	1	1	1	1	0
Phage phiN317	SA1770	hypothetical protein	1	1	1	1	1	0
Phage phiN318	SA1771	hypothetical protein	1	1	1	1	1	0
Phage phiN319	SA1772	hypothetical protein	1	1	1	1	1	0
Phage phiN320	SA1773	hypothetical protein	1	1	1	1	1	0
Phage phiN321	SA1774	hypothetical protein	1	1	1	1	1	0
Phage phiN322	SA1775	hypothetical protein,similar to scaffolding protein	1	1		1	1	0
Phage phiN323	SA1776	hypothetical protein	1	1	1	1	1	0
Phage phiN324	SA1777	hypothetical protein	1	1	1	1	1	0
Phage phiN325	SA1778	hypothetical protein	1	1	1	1	1	0
Phage phiN326	SA1804	hypothetical transcriptional regulator	0	0	0	0	1	1
Phage phiN327	SA1805	repressor homolog	0	0	0	0	1	1
Phage phiN328	SA1806	probable ATP-dependent helicase	0	0	0	0	1	1
Transposon Tn554	SA1953:tnpC	transposition regulatory protein	1	0	0	1	0	0
Transposon Tn554	SA1954:tnpB	transposition regulatory protein	1	0	0	1	0	0
Transposon Tn554	SA1955:tnpA	transposition regulatory protein	1	0	0	1	0	0
SCCmec	USA62:arcB	ornithine carbamyltransferase	1	0	0	0	1	0
SCCmec	USA64:arcD	ornithine antiporter	0	0	0	0	1	0
SCCmec	USA65:arcA	arginine deiminase	1	1			1	
SCCmec	USA66:arcR	arginine repressor	0	0	0	0	1	0
SCCmec	USA74:opp3B	oligopeptide permease	0	0	0	0	1	0
SCCmec	USA75:opp3C	oligopeptide permease	0	0	0	0	1	0
SCCmec	USA78:copA	ATPase copper transport	0	0	0	0	1	0

HA CA

Table 4.16 CC22 HA- versus CA-strains. The differences between HA (EMRSA-15 strains) and CA strains of CC22 were related only to 2 loci, ΦSa2mw and ΦSa3mw. In the original annotation, the putative function for all but four of these genes was unknown.

Loci	Gene ID	Product Description	E15-B3 ST22-IV	E15-B3 ST22-IV	E15-B3 ST22-IV	ST22-IV CA
Phage phi Sa 2mw	MW1386	conserved hypothetical protein	1	1	1	0
Phage phi Sa 2mw	MW1390	subfamily M23B unassigned peptidases	1	1	1	0
Phage phi Sa 2mw	MW1393	conserved hypothetical protein	1	1	1	0
Phage phi Sa 2mw	MW1400	hypothetical protein	1	1	1	0
Phage phi Sa 2mw	MW1401	hypothetical protein	1	1	1	0
Phage phi Sa 2mw	MW1405	hypothetical protein	1	1	1	0
Phage phi Sa 2mw	MW1408	hypothetical protein	1	1	1	0
Phage phi Sa 3mw	MW1895	subfamily M23B unassigned peptidases	0	0	0	1
Phage phi Sa 3mw	MW1896	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1897	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1898	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1899	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1900	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1901	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1902	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1903	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1904	hypothetical protein [Capsid protein]	0	0	0	1
Phage phi Sa 3mw	MW1905	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1906	portal protein (phage portal protein)	0	0	0	1
Phage phi Sa 3mw	MW1907	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1908	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1909	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1910	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1911	hypothetical protein	0	0	0	1
Phage phi Sa 3mw	MW1912	hypothetical protein	0	0	0	1

HA CA

To summarise, with respect to patterns of genes in the mobile genetic elements of the HA and CA strains, variation in the array profiles was noted in two main loci, ΦSa2 and ΦSa3. Further analysis was carried out to determine whether this was a real association between genotype and phenotype.

A group of CA-strains of CC1 were compared, one MSSA strain and two MRSA strains (Table 4.17). Only two of the three CA-strains carried genes characteristic of the ΦSa2 locus, but all three strains carried genes similar to ΦSa3. These data suggest there was a stronger association between carriage of ΦSa3mw genes and CA-status. In this group (CA-strains of CC1), patterns of variability were also noted for the mobile genetic elements ΦN315 and Tn554 (unique to USA400). A cluster of genes unique to WA-MRSA (MW0040, 0042, 0043 and 0047; hypothetical proteins) were also noted.

To determine whether the presence of the contiguous genes of ΦSa3mw (seen in only the CA-strains) extended to the other strains in the collection, data were analysed for presence/absence of the genes at this locus (full dataset not shown). The results showed they were unique to the CA-strains (except for one HA-strain, discussed below). However, not all CA strains within the collection carried these genes (Table 4.18). That is, the genes are carried predominantly by CA-strains, but it is not absolutely carried by all CA-strains. Irish-1, a HA-MRSA strain carried a variant form of these genes in which most were present, except the MW1898-1900 group. Table 4.18 shows the results obtained for the CA strains (including the two CA MSSA strains NCTC 8325 and MSSA476).

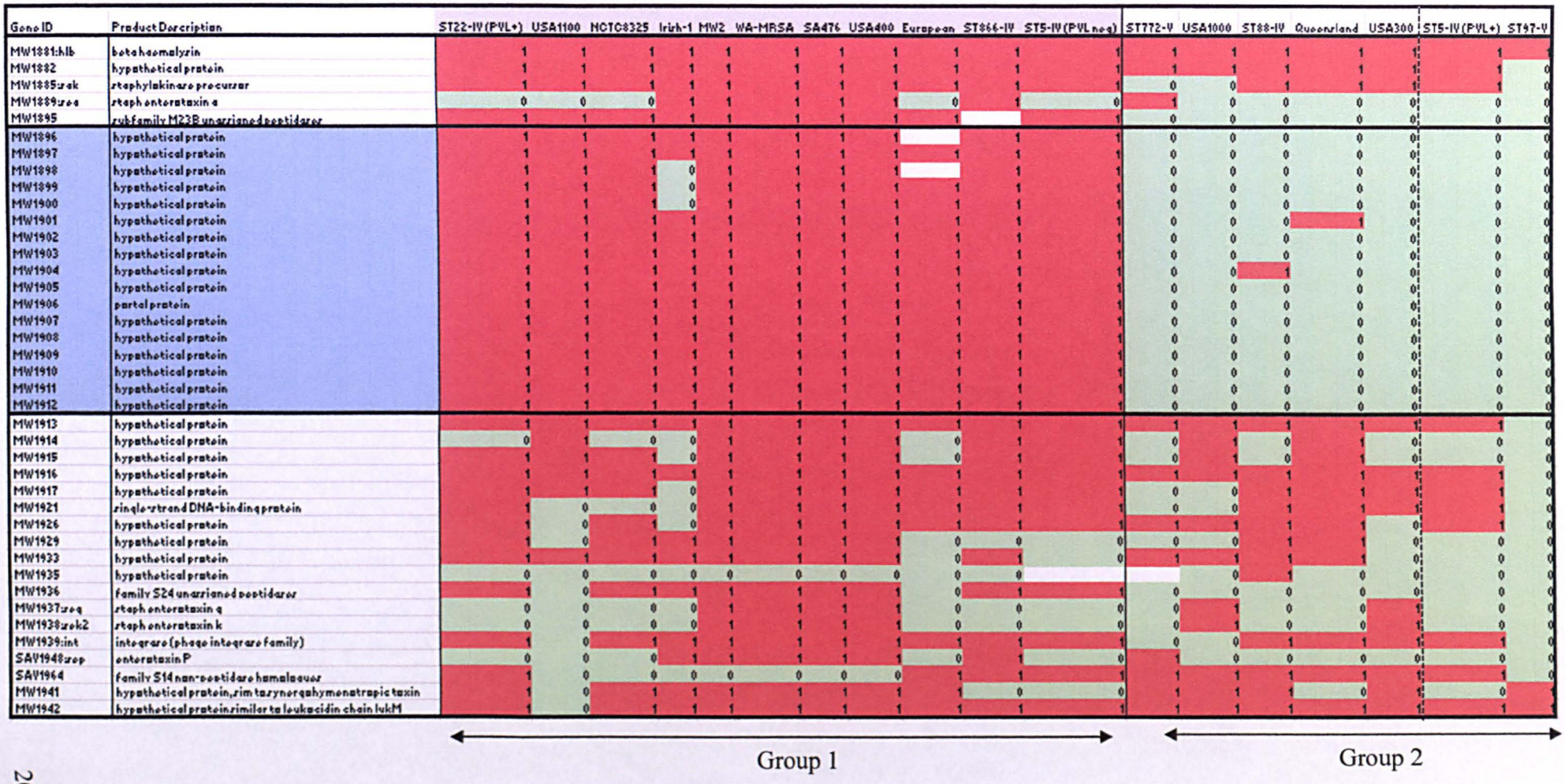
Table 4.17 CC1 Community associated strain comparison: MSSA v MRSA

The differences between CA-MSSA and MRSA of CC1 can be found on ΦSa2mw, ΦN315 and Tn554 loci. Variation in gene carriage is also observed for the block of genes in the region MW0040-47, which are located in the SCC*mec* element.

Loci	Gene ID	Product Description	SA476	WA-MRSA	USA400
SCCmec	MW0040	hypothetical protein	0	1	0
SCCmec	MW0042	hypothetical protein	0	1	0
SCCmec	MW0043	hypothetical protein	0	1	0
SCCmec	MW0045	hypothetical protein	0	1	1
SCCmec	MW0047	hypothetical protein	0	1	0
Bacteriophage phi Sa 2mw	MW1386	conserved hypothetical protein	1	0	1
Bacteriophage phi Sa 2mw	MW1390	subfamily M23B unassigned peptidases	1	0	1
Bacteriophage phi Sa 2mw	MW1393	conserved hypothetical protein	1	0	1
Bacteriophage phi Sa 2mw	MW1400	hypothetical protein	1	0	1
Bacteriophage phi Sa 2mw	MW1401	hypothetical protein	0	0	0
Bacteriophage phi Sa 2mw	MW1405	hypothetical protein	1	0	1
Bacteriophage phi Sa 2mw	MW1408	hypothetical protein	1	0	1
Bacteriophage phiN315	SA1808	probable ss-1,3-N-acetylglucosaminyltransferase	0	0	1
Bacteriophage phiN315	SA1809	hypothetical protein	0	0	1
Staphylococcus aureus transposon Tn554	SA1953:tnpC	transposition regulatory protein	0	0	1
Staphylococcus aureus transposon Tn554	SA1954:tnpB	transposition regulatory protein	0	0	1

MSSA MRSA

Table 4.18 ΦSa3mw CGH results for all CA strains analysed in this study (including Irish 1 – a HA-MRSA). Based on the genes MW1896-1912 (highlighted), the strains can be divided into two main groups: those carrying these genes (group 1; grey) and those without the genes (group 2). This pattern of presence/absence is not linked to PVL status, neither is it linked to strain lineage. The remaining genes on this phage do not show any particular pattern of association.



4.3.2 Putative functions of the ΦSa3 gene products

The ΦSa3mw genes (MW1896-1912) were mostly annotated as 'hypothetical proteins' at the original time of oligonucleotide design. To date, this has not been updated extensively, which prompted further *in silico* investigation. The array oligonucleotide sequences were compared with available annotations of all the sequenced strains from a local database. The BLAST utility was used to uncover similarities to genes with known function to provide insights into the putative functions of the prophage genes of interest. The results were selected by homology values, where high significance of homology was represented by low e-values. Putative functional assignments with respect to significant homologies of the ΦSa3mw genes (on the array) are listed in Table 4.19.

Of the genes of interest, the function of the original 'hypothetical proteins' could be predicted with high probability to functionally known genes only in 5 instances. These were the large terminase subunit (MW1909), the small terminase subunit (MW1908), the portal protein (MW1906), a prohead protease (MW1905) and the capsid protein (MW1904), that is, head and tail structural components involved in the phage lifecycle. The functions associated with these recognised genes are as follows. The phage terminase interact directly with the substrate DNA during DNA packaging into prohead shells (Bazinet and King, 1985). This interaction occurs at specified DNA interaction sites (*cos* and *pac* sites) which are generally located either within or close to the structural genes for the terminases (Black, 1989). During the packaging process, the portal protein associates with the terminase (Black, 1989).

Table 4.19 Putative ORFs for genes of ΦSa3mw (*in silico* analysis). Using a local BLAST utility, the genes for Sa3mw were blasted against all available sequenced annotations for *S. aureus* strains. Homologies to genes of known function have been listed. The morphology related genes have been highlighted.

Gene	Annotation in MW2	Related Proteins [Origin]	E value (% homol)
MW1883	Hypothetical protein	SAR2033 putative membrane protein[MRSA252]	2.00E-21 (100)
MW1885	Staphylokinase precursor (sak)	SAUSA300 1922 staphylokinase precursor (sak)[USA300-FPR3757]	4.00E-19 (98)
		SAR2039 staphylokinase precursor (sak)[MRSA252]	4.00E-19 (98)
MW1886	Initial annotation: HP	SAUSA300 1923 autolysin [USA300-FPR3757]	2.00E-21 (100)
	Update: lytic enzyme	SAR2040 autolysin [MRSA252]	2.00E-21 (100)
		SAOUHSC 02173 amidase[NCTC 8325]	2.00E-21 (100)
		NWMN 1881, note : amidase for bacteriophage phiNM3, phage amidase [Newman]	2.00E-21 (100)
		SaurJH1 2038 : CHAP domain containing protein [JH1]	2.00E-21 (100)
		SAOUHSC 02173 amidase[NCTC 8325]	2.00E-21 (100)
MW1887	Initial annotation: HP	SAUSA300 1924 holin[USA300-FPR3757]	2.00E-21 (100)
	Update: holin homolog	SAR2041 holin[MRSA252]	2.00E-21 (100)
		SAOUHSC 02174 holin, phage phi LC3 family[NCTC 8325]	2.00E-21 (100)
MW1889	Staphylococcal enterotoxin A precursor (sea)	MW1889 staphylococcal enterotoxin A precursor (sea)[MW2]	2.00E-21 (100)
		SAV1948 enterotoxin P (sep)[Mu50]	2.00E-21 (100)
		SAV1948 enterotoxin P (sep)[Mu50]	2.00E-21 (100)
		SAR2043 enterotoxin type A precursor[MRSA252]	4.00E-19 (98)
MW1890	Hypothetical protein	SAR2045 putative membrane protein[MRSA252]	2.00E-21 (100)
MW1892	Hypothetical protein	SAS061 hypothetical protein[N315]	2.00E-21 (100)
MW1893	Hypothetical protein	SAV1953 phi PVL ORF 20 and 21 homologue[Mu50]	2.00E-21 (100)
		SAUSA300 1928 phi77 ORF002-like protein, phage minor structural protein[USA300-FPR3757]	3.00E-04 (93.1)
		SAOUHSC 02180 phage minor structural protein, N-terminal region domain protein[NCTC 8325]	3.00E-04 (93.1)
MW1894	Hypothetical protein	SAUSA300 1929 phi77 ORF004-like protein, putative phage tail component[USA300-FPR3757]	2.00E-21 (100)
		SAOUHSC 02181 phi PVL orfs 18-19-like protein [NCTC 8325]	2.00E-21 (100)

		NWMN 1887, note : tail fiber protein for bacteriophage phiNM3, phage tail fiber protein [Newman]	2.00E-21 (100)
MW1895	Hypothetical protein	SAOUHSC 02182 tail length tape measure protein[NCTC 8325]	4.00E-19 (98)
MW1896	Hypothetical protein	SAOUHSC 02183 conserved hypothetical phage protein [NCTC 8325]	2.00E-21 (100)
MW1897	Hypothetical protein	SAOUHSC 02184 phi PVL orf 14-like protein[NCTC 8325]	2.00E-21 (100)
MW1898	Hypothetical protein	SAOUHSC 02185 phi PVL orf 13-like protein[NCTC 8325]	2.00E-21 (100)
MW1899	Hypothetical protein	SAOUHSC 02186 phi PVL orf 12-like protein[NCTC 8325]	2.00E-21 (100)
MW1900	Hypothetical protein	SAOUHSC 02187 phage protein, HK97 gp10 family[NCTC 8325]	2.00E-21 (100)
MW1901	Hypothetical protein	SAOUHSC 02188 phage head-tail adaptor, putative[NCTC 8325]	2.00E-21 (100)
MW1902	Hypothetical protein	SAOUHSC 02189 conserved hypothetical phage protein[NCTC 8325]	2.00E-21 (100)
MW1903	Hypothetical protein	SAOUHSC 02190 conserved hypothetical phage protein[NCTC 8325]	2.00E-21 (100)
MW1904	Capsid protein	SAOUHSC 02191 phage major capsid protein, HK97 family[NCTC 8325]	2.00E-21 (100)
MW1905	Hypothetical protein	SAOUHSC 02193 prohead protease[NCTC 8325]	2.00E-21 (100)
MW1906	Portal protein	SAOUHSC 02194 phage portal protein, HK97 family[NCTC 8325]	2.00E-21 (100)
MW1907	Hypothetical protein	SAOUHSC 02195 phi PVL orf 3-like protein-related protein[NCTC 8325]	4.00E-19 (98)
MW1908	Hypothetical protein	SAOUHSC 02196 phage terminase, large subunit, putative[NCTC 8325]	2.00E-21 (100)
MW1909	Hypothetical protein	SAOUHSC 02197 phage terminase, small subunit, putative[NCTC 8325]	2.00E-21 (100)
MW1910	Hypothetical protein	SAOUHSC 02198 conserved hypothetical phage protein[NCTC 8325]	2.00E-21 (100)
MW1911	Hypothetical protein	SAOUHSC 02199 phi PVL orf 62-like protein[NCTC 8325]	2.00E-21 (100)
MW1912	Hypothetical protein	SAOUHSC 02200 conserved hypothetical phage protein[NCTC 8325]	4.00E-07 (88)
MW1913	Hypothetical protein	SAUSA300 1945 phi77 ORF071-like protein[USA300-FPR3757]	7.00E-22 (100)
		SAR2068 putative exported protein[MRSA252]	7.00E-22 (100)
MW1914	Hypothetical protein	MW1914 hypothetical protein [MW2]	7.00E-22 (100)
MW1915	Hypothetical protein	MW1915 hypothetical protein [MW2]	7.00E-22 (100)
MW1916	Hypothetical protein	MW1916 hypothetical protein [MW2]	7.00E-22 (100)
MW1917	Hypothetical protein	SAUSA300 1954 phiPVL ORF050-like protein[USA300-FPR3757]	7.00E-22 (100)
		SAOUHSC 02211 phi PVL orf 50-like protein[NCTC 8325]	7.00E-22 (100)
		SAUSA300 1416 phiSLT ORF 81b-like protein [USA300-FPR3757]	4.00E-13(95.45)

MW1921	Single-strand DNA-binding protein	SAUSA300 1958 Single-strand binding protein [USA300-FPR3757]	7.00E-22 (100)
		SAR2083 putative single-strand DNA-binding protein [MRSA252]	7.00E-22 (100)
MW1923	Hypothetical protein	SAUSA300 1960 putative phage-related DNA recombination protein [USA300-FPR3757]	2.00E-21 (100)
MW1925	Hypothetical protein	SAR2087 hypothetical phage protein [MRSA252]	2.00E-14 (94)
MW1926	Hypothetical protein	SAOUHSC 02223 phi PVL orf 39-like protein[NCTC 8325]	7.00E-22 (100)
		SAUSA300 1427 phiSLT ORF86-like protein[USA300-FPR3757]	2.00E-07 (88)
MW1927	Hypothetical protein	SAOUHSC 02224 phi PVL orf 38-like protein-related protein [NCTC 8325]	2.00E-21 (100)
		SAR2091 putative exported protein [MRSA252]	4.00E-19 (98)
		SAUSA300 1429 phiSLT ORF53-like protein [USA300-FPR3757]	6.00E-12 (92)
MW1929	Hypothetical protein	SAOUHSC 02226 conserved hypothetical phage protein[NCTC 8325]	7.00E-22 (100)
MW1931	Hypothetical protein	SAR2095 hypothetical phage protein [MRSA252]	2.00E-21 (100)
MW1932	Initial annotation: HP	SAUSA300 1966 phi77 ORF014-like protein, phage anti-repressor protein [USA300-FPR3757]	2.00E-21 (100)
	Update: phage anti repressor	SAR2096 putative anti repressor [MRSA252]	2.00E-21 (100)
		SA1801 anti repressor [N315]	2.00E-21 (100)
MW1933	Hypothetical protein	SAOUHSC 02232 phi PVL orf 33-like protein[NCTC 8325]	7.00E-22 (100)
MW1934	Hypothetical protein	SAOUHSC 02233 phi PVL orf 32-like protein [NCTC 8325]	2.00E-21 (100)
MW1935	Hypothetical protein	SAOUHSC 02234 repressor-like protein-related protein[NCTC 8325]	7.00E-22 (100)
MW1936	Hypothetical protein	MW1936 phage repressor[MW2]	7.00E-22 (100)
		SACOL0321 prophage L54a, repressor protein, putative[COL]	2.00E-19 (98)
		SAOUHSC 02084 phage repressor protein, putative[NCTC 8325]	2.00E-19 (98)
MW1937	Staphylococcal enterotoxin SEG (seg2)	SACOL0887 staphylococcal enterotoxin type I (sei)[COL]	7.00E-22 (100)
MW1938	Staphylococcal enterotoxin Sek (sek2)	SAUSA300 0800 staphylococcal enterotoxin K (sek)[USA300-FPR3757]	2.00E-19 (98)
		SACOL0886 staphylococcal enterotoxin (sek)[COL]	2.00E-19 (98)
MW1939:int	Integrase (int)	SAUSA300 1972 integrase (int)[USA300-FPR3757]	7.00E-22 (100)
		SA1810 integrase (int) [N315]	7.00E-22 (100)
MW1941	HP, sim to synergohymenotropic toxin precursor	SAUSA300 1974 Leukocidin/Haemolysin toxin family protein[USA300-FPR3757]	7.00E-22 (100)
MW1942	HP, sim to leukocidin	SAUSA300 1975 Aerolysin/Leukocidin	7.00E-22 (100)

	chain lukM precursor	family protein[USA300-FPR3757]	
		SAB1876c probable leukocidin S subunit[RF122]	7.00E-22 (100)
		SA1813 hypothetical protein, similar to leukocidin chain lukM[N315]	7.00E-22 (100)
		SACOL2006 Aerolysin/Leukocidin family protein[COL]	7.00E-22 (100)

Many of the ΦSa3mw genes shared homology to ORFs of ΦPVL, ΦSLT and Φ77 (see Table 4.19). This prompted research into the area of staphylococcal temperate phage. Studies into staphylococcal genomes have been extensive and more than 80 phage genomes have been completely sequenced (Kaneko *et al.*1998, Iandolo *et al.* 2002, O'Flaherty *et al.* 2004, Kwon *et al.* 2005). Previous studies analysed the relatedness of *S. aureus* bacteriophage by means of techniques such as DNA hybridisation (Pariza and Iondolo 1974, Inglis *et al.* 1987, Stewart *et al.* 1985), high resolution thermal denaturation analysis of DNA (Inglis *et al.*, 1987), restriction endonuclease similarity patterns (Doskar *et al.*, 2000) and virion protein profiles (Lee and Stewart, 1985). These data show staphylococcal temperate phage can be classified into genotypically defined groups and that sequence similarities correlate with serological relatedness rather than host specificity (Pantucek *et al.*, 2004).

Comparative analysis of ΦSa3 and ΦPVL variants revealed a high degree of similarity over most of their genomes; however, differences in virulence gene content were noted, e.g. enterotoxins *seg2* and *sek2* found only in ΦSa3 (Brussow *et al.*, 2004). This was also observed in the *in silico* analysis for this study (data not shown), that showed many of the 'hypothetical genes' annotated as homologues of ΦPVL phage ORFs in strain NCTC 8325. Although, the ΦSa3 genomic islands are documented to be related to ΦPVL, only the latter group (ΦPVL phage) is associated with PVL carriage. One explanation for this relationship could be homologous recombination. The genes of ΦSa3 islands are notable for their mosaic structures, comprising not only ΦPVL genes but also elements from ΦSLT (another PVL-carrying phage). It has been proposed that the patchwork similarities (and therefore high diversity) between *S. aureus* phage genomes is a result of multiple recombination events and horizontal transfer between Φ N315-, ΦPVL-, ΦPV83-, ΦSa3-,

and Φ13-like phages during their evolutionary history (Brussow *et al.* 2004, Canchaya *et al.* 2003).

In general, phage genomes are modular. In members of the family *Siphoviridae* to which staphylococcal phage belong (figure 4.6), each module contains a set of genes which carry out a distinct biological function (Casjens *et al.*, 1992). Phage evolution proceeds via the exchange of modules from the same gene pool (Botstein 1980). This theory was defined by Botstein (1980) as the theory of modular evolution that proposes "*joint evolution of sets of functionally and genetically interchangeable elements each of which carries out a particular biological function*" may occur. Genetic exchange in this way has not only played a fundamental role in the evolution of the phages, but also in the bacterial host (Fitzgerald *et al.* 2001, Brussow *et al.* 2004). This molecular organisation and conservation of the order of genes is shared by many phages of the same genome-size class (Kwon *et al.*, 2005).

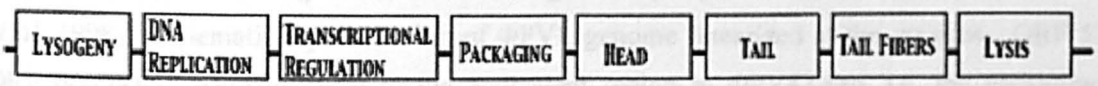


Figure 4.6 The conserved modular structure of staphylococcal phage

Conservation of the morphogenesis module (head and tail structural components) in some tailed phages was first pointed out in the early 1970s during the comparative analysis of the prophage maps of P2, P22 and λ which showed that they were partially congruent (Dove, 1971). This was later confirmed by sequence comparison of P22 and λ (Eppler *et al.*, 1991). Desiere and colleagues (1999) in their comparative analysis of the DNA packaging and structural modules of *Streptococcus thermophyilus* phage Sfi21, noted that during bioinformatics comparison with other sequenced strains, homologies generally corresponded to phage genes with corresponding positions on their respective genome

maps. This method of comparison they called comparative sequence 'gazing', and highlighted its ability to increase biological understanding of phage genomics (Desiere *et al.*, 1999). Therefore, to assign putative functions to the remaining hypothetical genes of the group of interest in this study, where possible they were mapped to genes in related staphylococcal bacteriophage (figure 4.7).

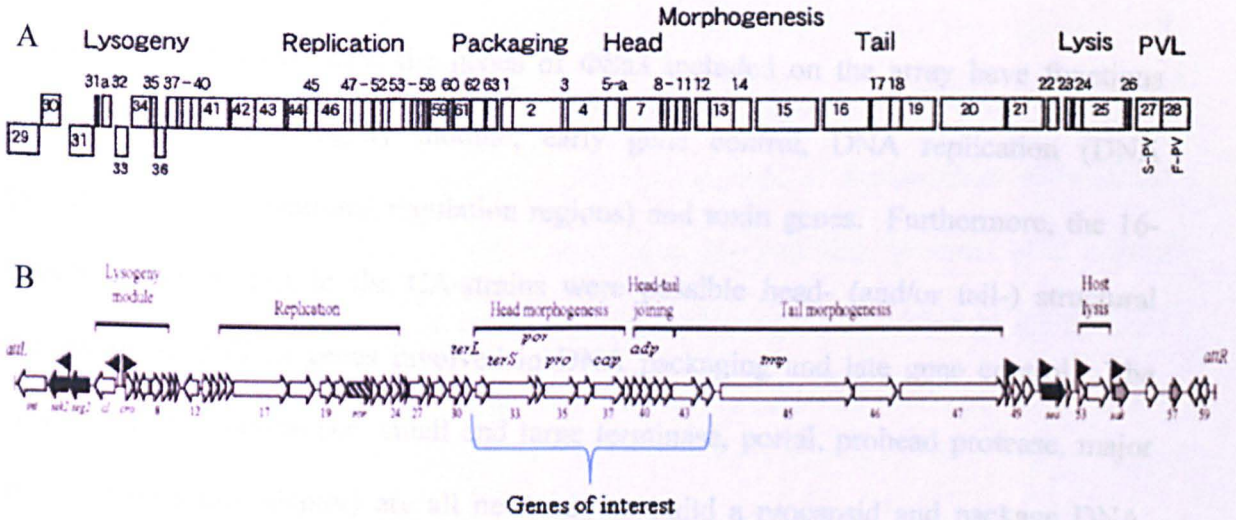


Figure 4.7 (A) Genetic organisation of ΦPVL genome (adapted from Zou *et al.* 2000 and Kaneko *et al.* 1998). Schematic representation of ΦPVL genome linearized at the *att* sites. ORF 57 and ORF 60-ORF22 were identical to the equivalent region in ΦPV83-pro, i.e. the packaging and morphogenesis region. (B) Genetic organisation of the ORF map of the ΦSa3ms genome.

A further classification system, the phage proteomic tree classifies phage into taxa based on the overall similarity of about 100 completely sequenced phage genomes (Rohwer and Edwards, 2002). It addresses phage taxonomy independently of phage phylogeny; the former resolves around the naming although it does not need to deal with phylogeny. "*The tree is the result of a computer-based (objective) calculation process and is applicable to any completely sequenced phage*". Using this technique, ΦSLT, Φ12, ΦPVL, Φ13, and ΦPV83 were all classified as PVL-like *Siphoviridae*. The confusion regarding the established and more recently proposed classification of phage taxonomy has been

recognised. Proux and colleagues (2002), and Nelson (2004) have both highlighted this dilemma. They concluded that "*the real taxonomic value of the different proposals will depend on the diversity of the bacteriophage genomes encountered in the environment*".

4.3.3 ΦSa3 locus

From the data in Table 4.19, the genes of ΦSa3 included on the array have functions associated with the lysogeny module, early gene control, DNA replication (DNA replication and transcriptional regulation regions) and toxin genes. Furthermore, the 16-consecutive genes seen in the CA-strains were possible head- (and/or tail-) structural components, as well as genes involved in DNA packaging and late gene control. The genes of known function (i.e. small and large terminase, portal, prohead protease, major capsid and head-tail adaptor) are all necessary to build a procapsid and package DNA. However, whether all the phages of the strains in this study have the necessary tail structural proteins to produce infectious particles, is not something that can be commented on from this array data set since the genes encoding these structures were not included on the array. Comparative sequence gazing revealed that the array genes of ΦSa2 and ΦSa3, which are mostly components of morphology related modules, varied in their presence/absence in this group of strains. This implies that there may be different morphotypes of these phages circulating within the staphylococcal population.

DNA packaging mechanisms can be differentiated into two groups i.e. those used by cos site and pac site phages (Le Marrec *et al.*, 1997). Research into the DNA packaging mechanisms of the Sfi21 (cos) and Sfi11 (pac) phages of the *Siphoviridae* family revealed

that two clearly distinct structural gene clusters corresponded to these packaging mechanisms (Lucchini *et al.*, 1998). This led to division of the genus *Siphoviridae* into two groups for the Sfi21-like and Sfi11-like phage. Comparative genomics was used to find these phage groups in many bacterial genera (Desiere *et al.*, 2001). Conservation of head and/or tail structural proteins has been used as a method for assigning phages to lineages. Classification of phage genomes based on structural genes has been shown to be a discriminatory technique to distinguish bacterial sub species. For example comparative genomics of the structural genes of lactococcal phages delineated four species and distinguished two genera based on head morphotype (Proux *et al.*, 2002). Similarly, Brussow and Desiere (2001) in their comparative analysis of *Siphoviridae* showed the relatedness and conserved order of the DNA packaging and head proteins of *S. aureus* phages ΦSLT, ΦPVL and ΦPV83. The evolutionary history of individual phage modules (e.g. this structural module) is starting to emerge through such comparative studies (Brussow and Desiere, 2001).

Narita and colleagues (2001) compared the genomes of three PVL-carrying phages, ΦSLT, ΦPVL and ΦPV83-pro. The mosaic structure of these phage was shown. For example, ΦPV83-pro was shown to have the *att* site and integrase of Φ11 together with packaging and morphogenesis modules from ΦPVL. Thus it seems likely that the ΦSa3 phage are mosaics comprised of elements found in other staphylococcal phages.

4.3.4 Prophinder software: comparative analysis of staphylococcal prophage genomes

Genomic analyses of the prophage content of ΦSa3mw (MW2), ΦSa3ms (MSSA476) have shown that only 14bp difference exists between the sequences (Sumby and Waldor, 2003). To investigate the structural relationships between the morphogenesis units of phages found in completely sequenced strains that carried the genes of interest (e.g. MW2, MSSA476 and NCTC8325) and those that do not (e.g. USA300) the Prophinder online tool (http://aclame.ulb.ac.be/Tools/Prophinder/) was used. Prophinder detects prophages in sequenced bacterial genomes using BLASTP. This algorithm detects coding sequences that are similar to phage proteins stored in the local ACLAME database. Similarities were found with six phage i.e. Φ13, ΦPVL, ΦPV83, Φ108-PVL, tp310-1 and tp310-3 (spa-310 type staphylococcal phages from a *pvl*$^{+}$ strain, MLST22). The region of significant homology was present in ΦSa3mw. In comparison, strain USA300 (which did not harbour the block of genes putatively associated with the CA phenotype) showed most homology to ΦNM3 (from Newman strain), ΦN315 and Φ77 within this locus. This is further evidence that the prophage of NCTC 8325, MSSA 476, and MW2 are related and distinct from the prophage of USA300. It also supports the suggestion the eleven CA-strains that carry the specific head and tail gene unit (Table 4.18) would also show homology to the same six phages. Conversely, it appears that the phage of USA300, and other strains that do not carry this block of genes, are of a different morphotype.

Although there are multiple proposed theories for lineage determination, the following may be speculated with respect to these studies. If the head structural components are conserved in a proportion of the CA-strains, but are not seen in the hospital strains (except a variant in the Irish-1 strain), it may be possible that a different 'type' of phage is

associated with this group, and therefore associated with community-acquiredness. It could also be speculated that the proposed variant phage is associated with the MSSA strains which have gone on to become successful CA-strains.

A recent study by Ma and co-workers (2008) presented data on the prevalence of PVL phage in MRSA and MSSA strains in Japan. In this study, the entire nucleotide sequences of two bacteriophage carrying PVL (ΦSa2958 and Φ108-PVL) were compared. Both phage integrate at the site corresponding to ΦSa2mw on the MW2 genome. By comparison to the extant five PVL phages characterised to date (ΦPVL, ΦSLT, ΦSa2mw, Φ108PVL and ΦSa2usa), ΦSa2958 displayed conservation over 27 open reading frames (ORF) with the phages ΦSa2mw and ΦSLT. In particular, this included 8 ORFs associated with DNA packaging and head and tail formation (i.e. the terminase small subunit, terminase large subunit, portal protein, prohead protease, capsid protein, major tail protein and phage tail tape measure proteins). The phage morphologies of the five characterised phages are not identical, but can be assigned to two types: the icosohedral-head type and the elongated-head type (Canchaya *et al.*, 2003). In these strain populations they observed both the icosohedral-head type (represented by similarity to Φ108-PVL) and the elongated head type phages (represented by similarity to ΦSa2958). Ma and colleagues observed no link between SCC*mec* type and carriage of a particular PVL phage type. However, they did observe that most of the PVL positive strains were of sequence type ST30 (or belonged to CC30), and produced coagulase type 4. It has therefore been proposed that these two groups of phages (or similar phages) were independently acquired by MSSA strains (coagulase type 4 ST30) prior to the acquisition of SCC*mec* elements, and later evolved into PVL-positive MRSA strains (Ma *et al.*, 2008). Ma and co-workers concluded that phages ΦSa2958, ΦSa2mw and ΦSLT belong to a taxon that is associated with the virulent

PVL^+ phenotype in Japan (Ma *et al.*, 2008). The data presented in this study support these findings i.e. that a particular phage taxon, in this case ΦSa3mw, is also associated with the CA phenotype.

Phage head conservation is also deduced from the array data available for the genes of bacteriophage ΦN315 and ΦSa2 (see the following sections). No particular distinction between HA- and CA-strains with respect to the pattern of these genes was identified. However, it was noted was that CA-strains carrying ΦSa3 DNA packaging/ morphogenesis genes, usually lacked the morphogenesis genes of ΦN315 (Table 4.20).

4.3.5 ΦN315 locus

In strain N315, bacteriophage ΦN315 integrates at a locus adjacent to ΦSa3 of strain MW2. Interestingly, the ΦN315 genes were present in the majority of strains tested (HA and CA) with the exception of seven CA strains and Irish-1 (Table 4.20); these strains all carried the ΦSa3mw gene block or the close variant found in Irish-1. Three isolates (CA-strains of STs 22, 866 and 5 [PVL neg]) carried both ΦSa3mw and ΦN315 gene blocks.

Table 4.20 ΦN315 CGH profile for all the CA-strains (and Irish-1). Most of the group-1 strains (carrying ΦSa3mw morphology genes) lack the equivalent morphology genes in ΦN315; the opposite is observed for the group-2 strains (most carry ΦN315 genes).

Gene ID	Product Description	ST22-IV(PVL+)	USA1100	NCTC8325	Irish-1	MW2	WA-MRSA	SA476	USA400	European	ST866-IV	ST5-IV(PVL neg)	ST772-V	USA1000	ST88-IV	Queensland	USA300	ST5-IV(PVL+)	ST97-V
SA1755	chemotaxis inhibitory protein	1	1	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0	0
SA1766	hypothetical protein	1	1	0	0	0	0	0	0	0	0	1	0		1	0	1	0	0
SA1768	hypothetical protein	1	1	0	0	0	0		0	0	1	1	1	1	1	1	1	1	0
SA1769	hypothetical protein	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1770	hypothetical protein	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1771	hypothetical protein	1	0	0		0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1772	hypothetical protein	1	0	0	0	0	0	0		0	1	1	1	1	1	1	1	1	0
SA1773	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1774	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1775	hypothetical protein, similar to scaffolding protein	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0
SA1776	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1777	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
SA1778	hypothetical protein	1	0	0	0	0	0		0	0	1	1	1	1	1	1	1	1	0
SA1791	hypothetical protein	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	0
SA1794	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SA1795	hypothetical protein	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	0
SA1797	hypothetical protein	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0
SA1800	hypothetical protein	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
SA1804	hypothetical transcriptional regulator	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
SA1805	repressor homolog	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
SA1806	probable ATP-dependent helicase	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
SA1808	probable ss-1,3-N-acetylglucosaminyltransferase	0	0	0	0	0	0	0	1	0	0	0		0	0	0	0	0	0
SA1809	hypothetical protein	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Group 1 Group 2

4.3.6 Bioinformatics analysis of ΦN315 gene products

Most of the ΦN315 genes on the array were originally annotated as 'hypothetical protein'. Functional assignment of these genes were explored using the methods described previously, revealing 11 consecutive ORFs involved in head morphogenesis. Thus it was possible to identify the corresponding morphology genes of ΦSa3 and ΦN315 (Table 4.20 and Table 4.21).

Several phage integrate into the staphylococcal genome at positions adjacent to ΦSa3mw (e.g. ΦSa3ms). These have been reported to display sequence similarities (Sumby and Waldor, 2003). However, in most cases the strains tested either had homologies to ΦSa3mw or to ΦN315 with little evidence for mosaic structures. As discussed above Prophinder analysis showed homology between the prophages ΦN315, ΦNM3 and Φ77. This group is distinct from that of ΦSa3mw i.e. the Φ13, ΦPVL, ΦPV83, Φ108-PVL, tp310-1 and tp310-3 group. Yet, interestingly, CA strains ST22, ST866 and ST5 (PVL neg) were unusual in that they carried the morphogenesis genes from both ΦSa3mw and ΦN315. It is unclear whether these strains carry a phage intermediate between ΦSa3mw and ΦN315 or if both are present. Finally a single unusual CA-strain (ST97 SCC*mec* type V) carried neither set of structural genes.

Of the HA-MRSA strain collection, four strains did not have structural genes corresponding to ΦN315, these were strains COL, Iberian ST247, EMRSA-15 variant B3 and EMRSA-15 variant B5. Additionally, the animal associated CA-MRSA strain ST398,

did not carry these genes. The COL genome (which has been fully sequenced) does not have any prophage sequences at this locus (data from Prophinder). Therefore, it may be that the other 4 HA-MRSA strains also do not have prophage at this locus. However, this cannot be confirmed here. A limitation of the array is that dissimilarity to a present prophage and complete absence of a prophage cannot be distinguished.

Table 4.21 Putative ORFs for genes of ΦN315 (*in silico* analysis). Using a local BLAST utility, the genes for ΦN315 were blasted against all available sequenced annotations for *S. aureus* strains. Homologies to genes of known function have been listed. The morphology related genes have been highlighted.

Gene	Annotation in N315	Related Proteins [Origin]	E value (% homol)
SA1755	Hypothetical Protein	SAUSA300 1920 chemotaxis-inhibiting protein CHIPS (chs) [USA300-FPR3757]	3.00E-06 (100)
		SAR2036 chemotaxis-inhibiting protein CHIPS (chp) [MRSA252]	3.00E-06 (100)
SA1759	Hypothetical Protein	SAUSA300 1923 autolysin [USA300-FPR3757]	1.00E-16 (100)
	Update: lytic enzyme	SAR2040 autolysin [MRSA252]	1.00E-16 (100)
		MW1886 lytic enzyme [MW2]	1.00E-16 (100)
		SAOUHSC 02173 amidase [NCTC 8325]	1.00E-16 (100)
		NWMN 1881, phage amidase [Newman]	1.00E-16 (100)
		SAB0781 lytic enzyme [RF122]	6.00E-12 (95.24)
SA1761	enterotoxin P (sep)	MW1889 staphylococcal enterotoxin A precursor (sea) [MW2]	9.00E-05 (86)
		SAV1948 enterotoxin P (sep) [Mu50]	9.00E-05 (86)
SA1766	Hypothetical Protein	SAR2050 putative membrane protein [MRSA252]	7.00E-22 (100)
		SAUSA300 1930 phi77 ORF001-like protein, phage tail tape measure protein [USA300-FPR3757]	2.00E-19 (98)
		SAV1955 phi PVL ORF 15 and 16 homologue [Mu50]	2.00E-19 (98)
SA1768	Hypothetical Protein	NWMN 1890 : phage major tail protein [Newman]	2.00E-21 (100)
		SaurJH1 2050 : phage major tail protein, phi13 family[JH1]	2.00E-21 (100)
		SAUSA300 1934 phi77 ORF020-like protein, phage major tail protein [USA300-FPR3757]	4.00E-19 (98)
SA1769	Hypothetical Protein	SAR2055 hypothetical phage protein [MRSA252]	7.00E-22 (100)
SA1770	Hypothetical Protein	SAUSA300 1935 phi77 ORF029-like protein [USA300-FPR3757]	4.00E-17 (96)
SA1771	Hypothetical Protein	SAUSA300 1936 conserved hypothetical phage protein [USA300-FPR3757]	7.00E-22 (100)
SA1772	Hypothetical Protein	SAR2059 hypothetical phage protein [MRSA252]	7.00E-22 (100)
SA1773	Hypothetical Protein	SAUSA300 1937 phi77 ORF045-like protein [USA300-FPR3757]	7.00E-22 (100)
SA1774	Hypothetical Protein	SAUSA300 1938 phi77 ORF006-like protein, putative capsid protein [USA300-FPR3757]	7.00E-22 (100)

SA1775	Hypothetical Protein sim to scaffolding protein	SAR2062 putative Clp protease [MRSA252]	2.00E-21 (100)
		NWMN 1897 : phage Clp-like protease [Newman]	2.00E-21 (100)
		SAUSA300 1939 phi77 ORF015-like protein, putative protease [USA300-FPR3757]	4.00E-19 (98)
SA1776	Hypothetical Protein	SAUSA300 1940 phage portal protein [USA300-FPR3757]	1.00E-14 (100)
		SAR2063 hypothetical phage protein [MRSA252]	3.00E-12 (97.37)
SA1777	Hypothetical Protein	SAUSA300 1941 phi77 ORF003-like protein, phage terminase, large subunit [USA300-FPR3757]	7.00E-22 (100)
SA1778	Hypothetical Protein	SAUSA300 1942 conserved hypothetical phage protein [USA300-FPR3757]	7.00E-22 (100)
SA1779	Hypothetical Protein	SAUSA300 1943 phi77 ORF040-like protein [USA300-FPR3757]	2.00E-21 (100)
		NWMN 1901 :phage HNH endonuclease [Newman]	2.00E-21 (100)
		SaurJH1 2062 : HNH endonuclease[JH1]	2.00E-21 (100)
SA1781	Hypothetical Protein	SAUSA300 1945 phi77 ORF071-like protein [USA300-FPR3757]	2.00E-14 (96)
		SAR2068 putative exported protein [MRSA252]	2.00E-14 (96)
SA1783	Hypothetical Protein	MW1414 hypothetical protein [MW2]	4.00E-19 (98)
SA1785	Hypothetical Protein	SAOUHSC 02059 phi PVL orf 52-like protein [NCTC 8325]	4.00E-19 (98)
		SAR2074 hypothetical phage protein [MRSA252]	2.00E-14 (96)
SA1788	Hypothetical Protein	SaurJH1 2072 : PVL ORF-50 family protein[JH1]	2.00E-21 (100)
		SAUSA300 1421 phiSLT ORF122-like protein, DNA polymerase [USA300-FPR3757]	6.00E-15 (100)
		SAR1539 putative DNA-binding protein [MRSA252]	6.00E-15 (100)
SA1791	Hypothetical Protein	SAUSA300 1957 phiPVL ORF046-like protein [USA300-FPR3757]	7.00E-22 (100)
		SAR2082 putative phage regulatory protein (pseudogene) [MRSA252]	7.00E-22 (100)
SA1792	single-strand DNA-binding protein	SaurJH9 0322 : single-strand binding protein [JH9]	2.00E-21 (100)
		MW1921 single-strand DNA-binding protein [MW2]	4.00E-19 (98)
SA1794	Hypothetical Protein	SaurJH9 0320 : RecT protein, product : RecT protein, from JH9, complete genome.	2.00E-21 (100)
		SAUSA300 1960 putative phage-related DNA recombination protein [USA300-FPR3757]	1.00E-16 (96)
SA1795	Hypothetical Protein	SAUSA300 1961 phiPVL ORF41-like protein [USA300-FPR3757]	1.00E-19 (98)
SA1797	Hypothetical	SAUSA300 1427 phiSLT ORF86-like protein	3.00E-20

	Protein	[USA300-FPR3757]	(100)
		SAOUHSC 02074 phi PVL orf 39-like protein [NCTC 8325]	4.00E-19 (98)
		SAB1748c hypothetical phage-related protein [RF122]	6.00E-18 (97.92)
SA1799	Hypothetical Protein	SaurJH1 2084 : protein of unknown function [JH1]	2.00E-21 (100)
SA1800		SA1800 hypothetical protein [N315]	7.00E-22 (100)
SA1802	Hypothetical Protein	SAUSA300 1967 conserved hypothetical phage protein [USA300-FPR3757]	2.00E-21 (100)
		NWMN 1919 : conserved hypothetical protein [Newman]	2.00E-21 (100)
SA1804	Hypothetical Protein	SAUSA300 1968 putative phage transcriptional regulator [USA300-FPR3757]	7.00E-22 (100)
	Update: hypothetical transcriptional regulator	SAR2099 DNA-binding protein [MRSA252]	7.00E-22 (100)
		SAOUHSC 01574 Helix-turn-helix domain protein [NCTC 8325]	7.00E-22 (100)
SA1805	Hypothetical Protein	SAUSA300 1969 phi77 ORF011-like protein, phage transcriptional repressor [USA300-FPR3757]	2.00E-21 (100)
	Update: repressor homolog	SAR2100 putative repressor [MRSA252]	2.00E-21 (100)
		SAOUHSC 01575 Helix-turn-helix domain protein [NCTC 8325]	2.00E-21 (100)
		NWMN 1921 :phage cI-like repressor [Newman]	2.00E-21 (100)
SA1806	Hypothetical Protein	SAUSA300 1970 putative exonuclease [USA300-FPR3757]	2.00E-21 (100)
	Update: probable ATP-dependent helicase	SAOUHSC 01576 exonuclease family [NCTC 8325]	2.00E-21 (100)
		NWMN 1922 : phage exonuclease [Newman]	2.00E-21 (100)

4.3.7 ΦSa2 locus

Three main patterns of the ΦSa2 related genes were observed amongst the strains analysed (Table 4.22, top segment). In the first group all the genes were absent, the second group displayed carriage of all the genes and the final group carried all but the MW1401 gene encoding the terminase large subunit. The ORFs preceding this gene encode proteins involved in head and tail formation. Therefore, it seems likely that at least three distinct types of phage are represented at this locus. However, no particular pattern was seen with respect to the CA-phenotype.

It was noted that although most of the strains were PVL^+, this was not an absolute correlation with ΦSa2mw, a recognised PVL-carrying phage. This was also noted by Lindsay and colleagues (2006) in their analysis of PVL positive clinical isolates. Of their strain collection, only one isolate carried ΦSa2. This is unsurprising since the PVL genes (i.e. *lukF* and *lukS*) may be horizontally transferred between lineages. The degree of correlation between PVL and the ΦSa3 morphogenesis genes (described in this thesis) was similar to that between PVL and ΦSa2. Some PVL^+ (and PVL^-) CA-MRSA carried sequences homologous to different ΦSa3 genes. The ΦSa3 morphogenesis gene cluster seems therefore to represent a useful marker for the community-associated phenotype. Although there is currently no evidence for a causative link, Ma and colleagues (2008) noted possible associations between phage head morphology and virulence. It may be speculated that phage factors are the key to variations in epidemiology and pathogenicity noted between HA- and CA-strains. Table 4.22 illustrates the genotypes of the CA strains for the ΦSa2, ΦSa3 and ΦN315 morphology related genes.

Table 4.22 Comparison of the morphology related genes of ΦSa2mw, ΦSa3mw and ΦN315 loci for all the CA-strains (and Irish-1). The general relationship between carriage of ΦN315 and ΦSa3mw morphology genes is evident; that is, presence of one group, with absence of the other. The exceptions to this rule are the three isolates ST22, ST5-IV (PVL neg) and ST866. There is no particular link with the ΦSa2mw genes.

Loci	Gene ID	Product Description	ST22-IV (PVL+)	USA1100	NCTC8325	Irish-1	MW2	WA-MRSA	SA476	USA400	European	ST5-IV (PVL neg)	ST866-IV	ST772-V	USA1000	ST88-IV	Queensland	USA300	ST97-V
PhagephiSa2mw	MW1386	conserved hypothetical protein	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiSa2mw	MW1390	subfamily M23B unassigned peptidases	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiSa2mw	MW1393	conserved hypothetical protein	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiSa2mw	MW1400	hypothetical protein	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiSa2mw	MW1401	hypothetical protein	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0
PhagephiSa2mw	MW1405	hypothetical protein	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiSa2mw	MW1408	hypothetical protein	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0
PhagephiN315	SA1768	hypothetical protein	1	1	0	0	0	0		0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1769	hypothetical protein	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1770	hypothetical protein	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1771	hypothetical protein	1	0	0		0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1772	hypothetical protein	1	0	0	0	0	0	0		0	1	1	1	1	1	1	1	0
PhagephiN315	SA1773	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1774	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1775	hypothetical protein, similar to scaffolding protein	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1776	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1777	hypothetical protein	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
PhagephiN315	SA1778	hypothetical protein	1	0	0	0	0	0		0	0	1	1	1	1	1	1	1	0
PhagephiSa3mw	MW1896	hypothetical protein	1	1	1	1	1	1	1	1		1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1897	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1898	hypothetical protein	1	1	1	0	1	1	1	1		1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1899	hypothetical protein	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1900	hypothetical protein	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1901	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0
PhagephiSa3mw	MW1902	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1903	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1904	hypothetical protein [Capsid protein]	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0
PhagephiSa3mw	MW1905	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1906	portal protein (phage portal protein)	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1907	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1908	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1909	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1910	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1911	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
PhagephiSa3mw	MW1912	hypothetical protein	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0

Group 1 Group 2

Sumby and Waldor (2003) analysed the transcription of the toxins of the staphylococcal phage ΦSa3ms in the hypervirulent CA-strain MSSA 476. They showed that prophage induction (by mitomycin C treatment) caused the up-regulation of the phage-encoded toxins *sak*, *sae*, *seg2* and *sek2* transcripts. However, the authors did not confirm that the upregulated transcripts were translated. Other studies have shown that prophage induction can occur *in vivo* (Broudy *et al.* 2001, Voyich *et al.* 2003, Wagner and Waldor 2001, Waldor and Mekalanos 1996). Thus it is possible that prophage induction initiated by host factors plays a role in pathogenesis (Sumby and Waldor 2003). Understanding the biology of prophage may provide insights into the roles of phage encoded virulence factors. Sequencing of further *S. aureus* strains and phage genomes will also aid this cause, as will experimentation in animal models.

Finally, Christianson and co-workers (2007) used comparative genomics to study Canadian epidemic lineages of MRSA by microarray analysis. In their comparison of 8 HA- and 2 CA-strains, they identified 1 ORF (of a possible 2,741 ORFs on the array) encoding a metallo-beta-lactamase family protein specific to CA isolates. Although not bacteriophage encoded, this ORF was localised to the J1 region of the SCC*mec* element. This was suggested as a putative marker for the differentiation of HA- and CA-MRSA isolates. However, they noted no defined virulence factors whose presence or absence could differentiate HA- and CA-MRSA definitively.

In summary, this analysis of an international panel of HA- and CA-MRSA has highlighted a novel putative marker for differentiating isolates in these two groups. The work contributes to a growing body of evidence pointing toward the possible role of prophage in the emergence of the CA phenotype.

4.3.8 The relationship between putative serotype and CA phenotype

In a recent study, Goerke and colleagues (2009) described the diversity of prophage in dominant *S. aureus* clonal lineages. Through analysis of polymorphisms within the integrase gene, they established a classification system for the *Siphoviridae* family of staphylococcal prophage (fig. 4.8). Virulence gene carriage and inferred serogroup type (A, B, Fa or Fb) for each of these phage were also described. Serogroup type was based on capsid, tail and tail appendix protein sequences (the F group were sub-divided since the DNA packaging, head and tail genes belonged to different modules). The work showed that prophage integrating at the Sa2 and Sa3 loci are not restricted to a particular morphotype (although the integrase gene units are homologous); serogroup A and Fb phage integrate at the Sa2 locus, whilst serogroup Fa and Fb integrate at Sa3. By extrapolation of this study, their work showed that the CA phenotype-associated genes described in this work (MW1896–MW1912), to which homology was assigned (via the Prophinder tool) to 6 *Siphoviridae* prophage (Φ13, Φtp310-3, ΦPVL, ΦPVL108, tp310-1, ΦPV83), belong exclusively to the Fb serogroup. This indicates that genes MW1896–MW1912 of the ΦSa3mw locus are markers for Fb serogroup phage. Similarly, based on the data presented by Goerke *et al.* (2009), the ΦN315 and ΦSa2mw genes identify the Fa and A serotypes respectively (fig. 4.9). The only morphotype not identified by the array data is the group B serotype carried by the sequenced Mu50 strain which, based on the array data, showed no structural units similar to the A, Fa or Fb serotypes. Using this information, the phage serotypes carried by the CA strains (including the HA strain Irish-1) and the sequenced HA strains analysed on the array are shown in Table 4.23. The assignment was based on the array data for those strains that have not been sequenced, whilst the data for the sequenced strains was obtained by Prophinder analysis and from the research by Goerke and colleagues (2009).

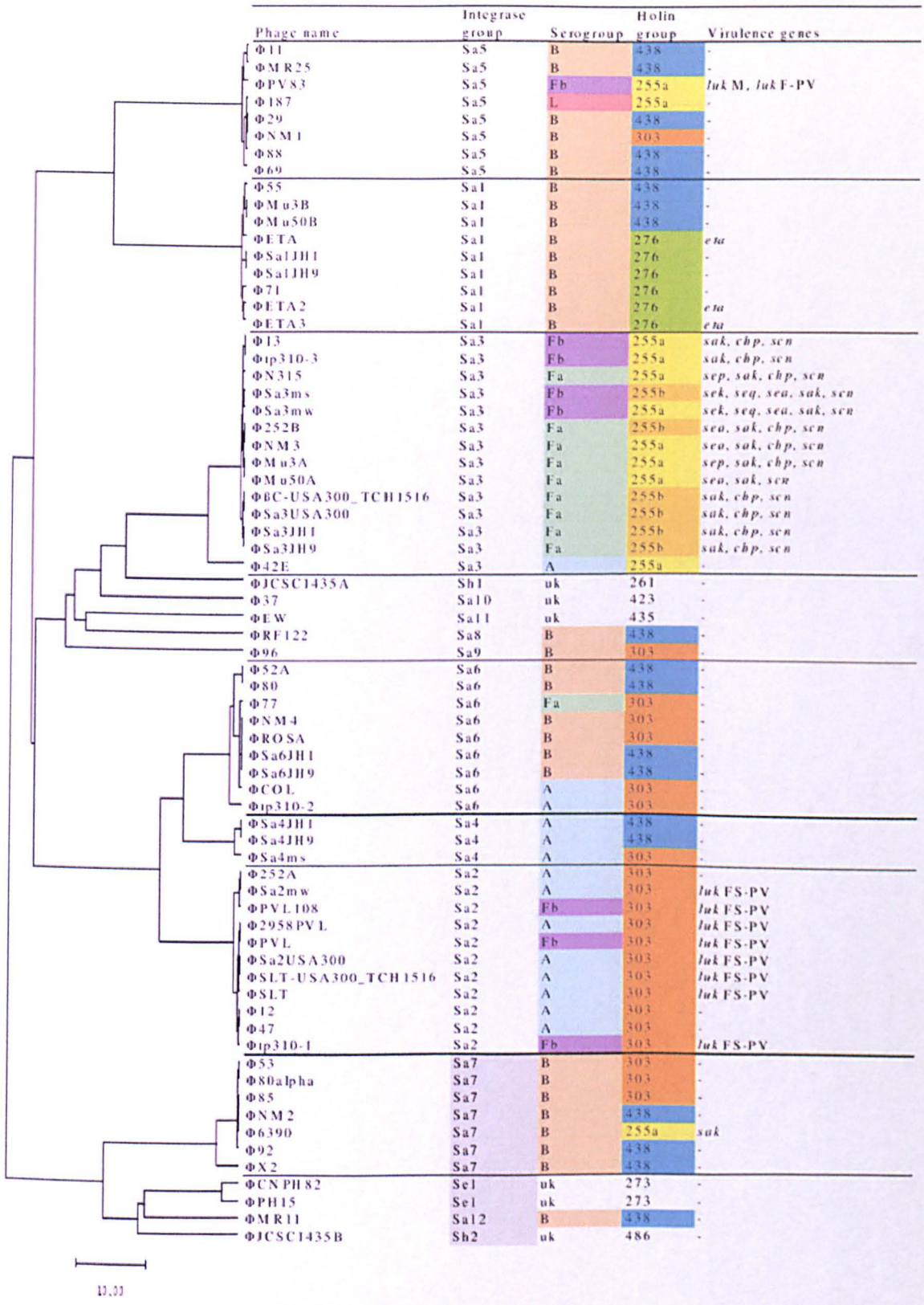


Figure 4.8 *In silico* analysis of the integrase groups, serogroups, holin groups, and virulence genes of 70 published staphylococcal bacteriophage of the class *Siphoviridae*. Integrase nucleotide sequences were aligned using the ClustalW algorithm. Identical serogroups and holing

groups are colour coded. Integrases of the serine recombinase-type family are shaded in grey. Taken from Goerke *et al.*, 2009.

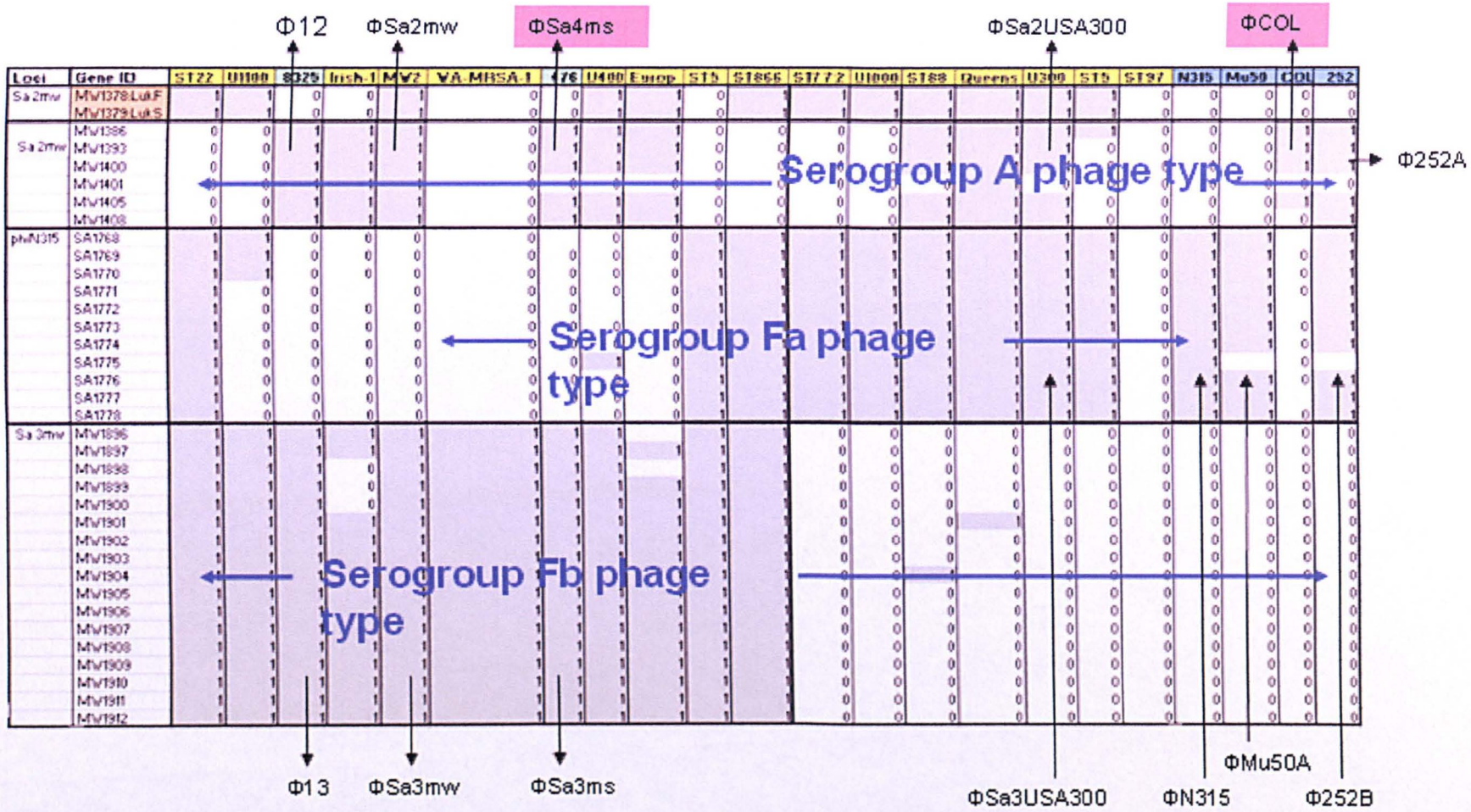


Figure 4.9 Comparisons of the morphology relates genes of ΦSa2mw (representing serogroup A phage type), ΦN315 (serogroup Fa phage type) and ΦSa3mw (serogroup Fb phage type) for all the CA strains analysed on the array. Irish-1, the HA strain that clustered with the CA strains is also included, as are the sequenced HA strains analysed on the array. CA-MRSA strains are highlighted in yellow, CA-MSSA in green, and HA-MRSA in blue (Irish-1 remains yellow). The first two rows indicate carriage of the *pvl* genes *lukFS*. The prophage data from Goerke *et al* (2009) has been mapped to this data, clearly indicating that the genes on the array correspond fully with phage serogroup type based solely on head/ tail sequences. No sites of integration can be specified from the array data, according to Goerke and colleagues (2009) ΦSa4ms and ΦCOL (highlighted in pink) integrate at different loci to the ΦSa2mw phage but all three are serogroup A phage.

Table 4.23 Putative relationship between *S. aureus* phenotype and bacteriophage serotype. Published sequenced strains are given in bold text. See text for details.

Strain name	HA/CA phenotype	Phage Serogroup at Sa3 locus or equivalent	Phage Serogroup at Sa2 locus or equivalent
ST22-IV (PVL+)	CA	Fa and Fb	np
USA1100 SWP: ST30-IV (PVL+)	CA	Part Fa and Fb	np
NCTC 8325: ST8 (MSSA)	CA	Fb	A
Irish-1: ST8-II (var)	HA	Variant of Fb	A
MW2 consensus	CA	Fb	A
WA-MRSA: ST1-IVa (PVL neg)	CA	Fb	np
SA476:ST1 (MSSA)	CA	Fb	A
USA400: ST1-IV (PVL+)	CA	Fb	A
European Clone: ST80-IV (PVL+)	CA	Fb	A
ST5-IV (PVL neg)	CA	Fa and Fb	np
ST866-IV (PVL+; tsst+)	CA	Fa and Fb	np
ST772-V (PVL+)	CA	Fa	np
USA1000 SE Asia: ST59-IV (PVL+)	CA	Fa	np
ST88-IV (PVL+)	CA	Fa	A
Queensland clone: ST93-IV (PVL+)	CA	Fa	A
USA300: ST8-IV (PVL+)	CA	Fa	A
ST5-IV (PVL+)	CA	Fa	np
ST97-V	CA	Fa	A
N315	HA	Fa	np
Mu50	HA	Fa	B
Mu3	HA	Fa	B
COL	HA	np	A
MRSA252	HA	Fa	A
Newman	HA	Fa	B
JH1	HA	Fa	B
JH9	HA	Fa	B

Group 1

Group 2

Sequenced strains

Carriage of *lukFS-pvl* is highlighted in Table 4.23 (grey; 'strain name' column). The data presented in the Goerke paper describes *pvl*-carrying phage to be either of the A or Fb serogroup. Those of the A serogroup (ΦSa2mw, Φ2958*PVL*, ΦSa2USA300, ΦSLT-USA300_TCH1516 and ΦSLT) and the majority of the Fb serogroup (Φ*PVL*, Φ*PVL*108 and Φtp310-1) integrate at the Sa2 position (the exception is ΦPV83 which integrates at the Sa5 locus). The data for the CA strains is summarised in Table 4.23. The vast majority

of the CA strains are *pvl*-carrying phage-positive with serogroup A and/or Fb phage present. Conversely, the HA strains predominantly carry Fa and B serogroup phage. Only in two cases (COL and MRSA252) were A serogroup phage noted, but neither was of the *pvl*-carrying type.

The absence of serogroup Fb phage from the HA strains implies that they may be involved in the CA phenotype. It was noted that the CA strains ST22-*pvl*$^+$ and USA1100 (ST30-IV *pvl*$^+$), both *pvl* positive, do not show presence of an A group phage. Therefore, *pvl* carriage in these cases is most likely associated with the Fb phage group. Additionally, with respect to WA-MRSA (ST1-IV *pvl*$^-$) and ST5-IV *pvl*$^-$ CA strains, both again did not show the presence of an A group phage; the Fb group was still present without the carriage of *pvl*. Since almost all the CA strains carry a PVL-associated phage, either of the A serogroup or the Fb serogroup (or possibly a mosaic phage within these divisions), it is believed that the carriage of PVL-associated phage, (but not *pvl* itself, since it is absent from some of the CA strains) is necessary for the CA phenotype. That is to say, the genes recognised by the array sometimes indicate the presence of an A/Fb serogroup phage carrying *pvl*, or in some cases, only a pvl-negative Fb phage. In the case of the latter, it is conceivable to believe that these strains may either carry a non-*pvl* associated Fb phage, or a mosaic phage of the Fb serotype, lacking the *pvl* genes. Fb phage could possibly form a mosaic phage (via exchange of modules according to the theory of modular evolution) that combines the fitness factors of the *pvl*-carrying Fb phage and the *sak/ chp/ scn* carrying Fb phage. Thus, to summarise, it seems that it could be the carriage of Fb- and/or A-serogroup prophage that are associated with CA-status. Since restriction modification play a central role in determining the uptake of bacteriophage by a bacterial genome, the serogroup A/Fb taxon phage associated with community-acquiredness may have a system by which this may be manipulated. The carriage of most of the morphology related genes

of the Fa and Fb serogroups by 3 of the CA strains, ST22 *pvl*$^{+}$, ST866 and ST5 *pvl* negative (confirmed for the latter two via sequence data; data not shown), provides supporting evidence for recombination between phage. Furthermore, detection of *pvl* phage not of the A/Fb groups (in strains ST772, USA1000 and ST5-IV *PVL*$^{+}$) indicates that *pvl* in this case could possibly be associated with a novel morphotype. According to the data by Goerke and colleagues, *pvl* is not associated with the Fa serogroup. That is to say, a *pvl* phage of a novel serotype could be circulating within CA strains.

The original identification of the unique morphology genes (now known to represent the Fb serogroup; MW1896– MW1912) in a subset of the CA strains described in this thesis prompted the sequencing of some of these members, including WA-MRSA, ST772, USA1000, ST866 and the Queensland clone ST93. Irish-1 was also sequenced to investigate its relationship with the CA strains (data not published). The sequenced data from these strains was used to confirm the presence of the phage serogroups described in Table 4.23 by nucleotide homology using BLAST analysis. Additionally, homology to the B serogroup (the only *Siphoviridae* serogroup not detected on the array) was analysed by nucleotide similarity to genes representing the structural units of φ11 (a group B phage). This *in silico* analysis showed that the data presented in Table 4.23 were correct, and that the B serogroup was not present in any of these CA strains. Additionally, Irish-1 did carry an Fb phage, not a variant as initially described, indicating that further analysis of this strain is required to fully understand its clustering with the CA strains. One possibility is that, like the other HA strains, a non-*pvl* type phage is carried by this strain (here a non-*pvl* Fb phage e.g. φ13, φtp310-3, φ Sa3mw or φSa3ms are possibilities). The evidence from this strain suggests that either this was an original CA strain that excelled in the hospital environment in which it was detected (where it became classified as a HA strain), or it shows the potential of HA strains to become CA-strains by acquiring additional factors.

The Sa3 prophage group, the largest group of *Siphoviridae* phage, are predominantly composed of serogroup Fa and Fb phage. Both these serogroups carry recognised virulence-associated genes other than *pvl*, including *sak*, *scn* and *chps* (a few also carry *sep/sea* or *seq*). From this there is no implication that the Fb phage are more virulent than the Fa group. However, it may be the combination of virulence-associated genes carried by the PVL-associated phage as well as the virulence-associated genes carried at the Sa3 position (*sak*, *scn*, *chps*, etc) produce a particularly virulent phenotype in those strains carrying phage of this morphotype. This putative link between a particular phage type and CA-status has not been previously reported. Analysis of further strains (in a revised version of the array) may provide insights into any further patterns between these genes and CA status. At present, these genes are a partial marker for the CA phenotype and in future it may be possible to establish a phenotype-genotype link. This finding further supports the role of the accessory genome, and in particular bacteriophage, in the evolution and epidemiology of *S. aureus*.

CHAPTER 5.0 TRANSCRIPTION PROFILING

The virulence-associated gene microarray was used to investigate gene expression in strains of interest. As an analytic technique, transcriptomic analysis provides more detailed information about bacterial response to its microenvironment. The results acquired purely reflect the transcripts obtained at the time of sampling, and thus the response to the experimental conditions at that time.

In this chapter, comparative analyses of an *agr* mutant strain are described with the aim of measuring changes in gene expression between a parental and mutant strain (see below). In addition, a *S. aureus* isolated from a chronic wound, grown under biofilm-simulating conditions, is analysed in the same way. Biofilm grown cells were compared to their planktonic counterparts to assess differences between expression levels, and also to determine the suitability of this model for further studies.

5.1 Agr study

5.1.1 Background

Agr, as described in the introduction (section 1.3.2.1), is one of the major and most thoroughly studied regulators of *S. aureus* virulence. A strain with a mutation in the *agr* locus (SH1001), and the progenitor unmodified strain (SH1000) were provided by Simon Foster (Horsburgh *et al.*, 2002). SH1000 was derived from the commonly used genetic lineage *S. aureus* 8325-4 (RN6390) and a functional *rsbU* gene (8325-4 *rsbU*$^{+}$) inserted. RsbU is the positive regulator of Sigma factor B (section 1.3.3), and is often non-functional in strain 8325-4 due to a mutation (8325-4 *rsbU*). For this study, strains were analysed for changes in expression levels at the early exponential- (5hr), late exponential-

(7.5hr) and stationary- (overnight, 21h) phases. The consensus model is that genes encoding surface-associated proteins are generally expressed in the early exponential phase during cell proliferation (e.g. fibronectin binding protein A), whilst secreted proteins (e.g. α-haemolysin) are repressed. Conversely, in the stationary phase the opposite regulation effect has been reported (fig. 5.1). That is, secreted proteins (e.g. proteases/toxins), which aid spread and colonization, are up-regulated during later stages.

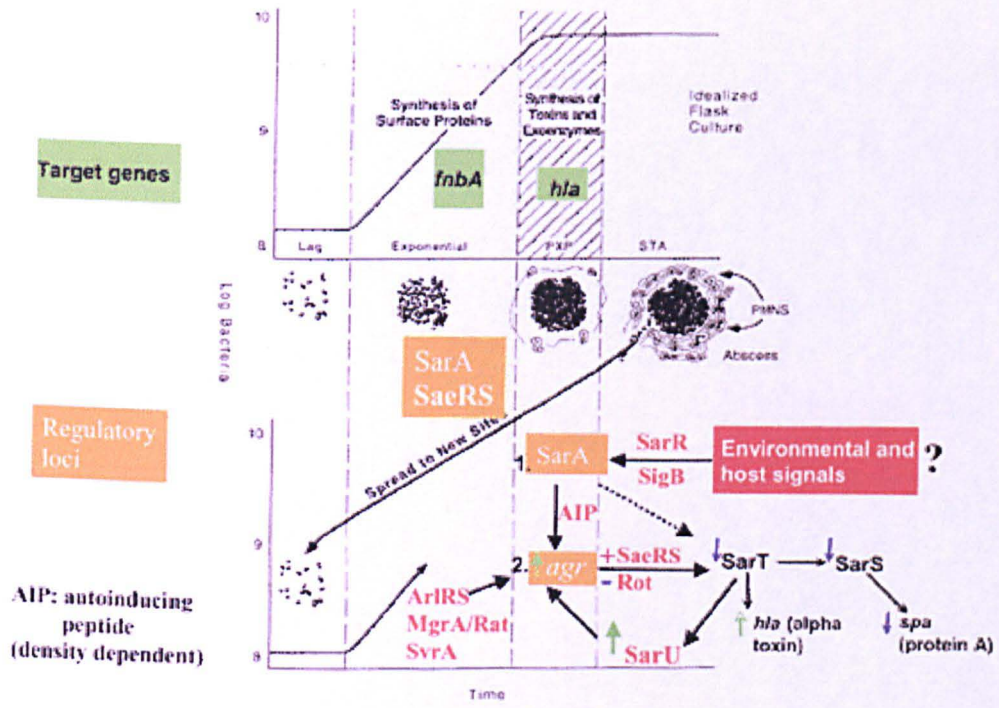


Figure 5.1 Growth phase dependent expression of virulence-associated genes observed during culture growth. Taken from Cheung *et al.*, 2004.

Cells were grown under the same conditions, on the same day, and extractions prepared in the same batch. The assumption is that genes that have changed response status (up or down regulation) between the isogenic strains could potentially be regulated via *agr*. The results obtained are described below.

5.1.2 *Agr*$^+$ strain

5.1.2.1 Differential expression observed between early exponential (EE) - versus late exponential (LE) -growth

In the *agr*$^+$ strain, comparison of EE- versus LE-phases revealed significant expression (>900 fluorescence units, average sum of Cy3 and Cy5 spot pixels) of 500 genes. Genes with weak signals in both channels were eliminated from this analysis. Logarithms (base 2) of the corrected ratios were recorded, and a crude cut-off threshold for significant expression change was set at +2 for down-regulated and -2 for up-regulated. In this study, it was often noted that individual gene transcripts were measured at a significant level at only one growth phase (EE). This indicates clear regulation of gene transcription, however the calculated ratio whilst clearly indicating up or down regulation, was not considered an accurate indication of the scale of regulation.

Based on these criteria, 13 genes were significantly down-regulated in the LE phase (Table 5.1). These genes were present above the detection threshold at one or both time points. Half of the 13 down-regulated genes in Table 5.1 were transporter proteins, mostly with iron related functions, that is, genes involved in metabolism. The LE phase represents the intermediate phase between actively growing cells and stationary cells. The down regulation observed probably reflects the fact that these metabolic genes are expressed less in the LE phase. Included also in this list were two transcriptional regulatory proteins, one of which is *gapR*, a protein involved in the glycolysis pathway for carbon metabolism. In addition, *spaA* (the IgG binding protein) was also significantly up-regulated (*spaA* is a gene known to be repressed by *agr*.

A greater number of genes (52) were found to be up-regulated between the EE and LE phases (Table 5.2). Several of these were related genes (some belonging to the same operon and others located next to each other around the chromosome), giving confidence in the data observed. These included the capsule genes (14 genes representing types 5 and 8 capsule groups), two serine proteases (MW1752-54) and several hypothetical proteins similar to (*i*) *Ear* protein – an exported protein (MW1757) and (*ii*) putative lantibiotic ABC transporters (MW1758-60) located adjacent to one another on genomic island nuSaβ2. Also up-regulated were members of the haemolysin toxin family (*hla*, *hla* precursor [SAR1136], *hlb*, *hlgB*, *hlgC*). The capsule genes represent a host defence mechanism, whilst the toxin and protease genes are also involved in the infection process. The lantibiotics protect the bacterium against self-produced toxins. As indicated previously the generally accepted model is that toxin and protease genes, which facilitate spread and colonization, are down-regulated in the early exponential phase and up-regulated post exponentially (fig. 5.1).

Table 5.1 Genes down-regulated in the *agr*$^+$strain between the EE and LE growth.

Genes with significant expression and $\log_2$ ratio > +2 are listed. TF (total fluorescence) units are given for each gene. Additional gene descriptions obtained from recent *in silico* investigations are given in square parentheses (for this table and all tables to follow).

Gene ID	Product description	$\log_2$ ratios (Agr^+ 5v7)	TF
MW2013	lipoprotein precursor	2.06	7177.52
SA1888: rodA	hypothetical protein, similar to rod shape determining protein RodA [cell division protein, FtsW/RodA/SpoVE family (ftsW)]	2.17	2779.52
MW0982:potA	spermidine/putrescine ABC transporter, ATP-binding protein homolog	2.50	1356.02
SA0598:pbp4	penicillin binding protein 4 (pbp4)	2.68	1232.02
SA2144	hypothetical protein, similar to transcriptional regulator (TetR/AcrR family)	2.78	8193.02
MW2103	hypothetical protein similar to ferrichrome ABC transporter (binding protein)	3.04	3040.52
MW0595	hypothetical protein, similar to ABC transporter ATP-binding protein	3.13	7902.52
SAV2455	family M42 unassigned peptidases (SAV2455 protein) [glutamyl-aminopeptidase]	3.28	4316.52
MW0084:spa	IgG binding protein A	3.46	39112.52
MW0594	conserved hypothetical protein [ABC transporter, permease protein]	3.54	7114.52
SA0726:gapR	glycolytic operon regulator [transcriptional regulator]	3.73	17808.52
MW0698	lipoprotein, similar to ferrichrome ABC transporter	3.94	2574.52
MW0593	lipoprotein, Streptococcal adhesin PsaA homologue [iron repressed lipoprotein (mntC)/ ABC transporter, substrate-binding protein]	4.54	11729.52

Table 5.2 Genes up-regulated in the *agr*$^+$ strain between the EE and LE growth.

Genes with significant levels of expression and $\log_2$ ratios < -2 are listed. Total fluorescence (TF) units are given for each gene.

Gene ID	Product description	$\log_2$ ratios (Agr$^+$ 5v7.5)	TF
SAV0161:Cap5M	capsular polysaccharide synthesis enzyme Cap5M	-2.02	20233.52
SAV0155:Cap5G	capsular polysaccharide synthesis enzyme Cap5G	-2.03	18497.52
MW1874	hypothetical protein, similar to ABC transporter, ATP-binding protein	-2.08	22357.52
MW1872	hypothetical protein, similar to ABC transporter, ATP-binding protein	-2.12	22189.52
SAV0154:Cap5F	capsular polysaccharide synthesis enzyme Cap5F	-2.14	17555.52
MW0764:clfA	fibrinogen-binding protein	-2.14	39294.52
MW1760:bsaF	hypothetical protein, similar to EpiF (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-2.19	13542.52
SAV0151:Cap8C	capsular polysaccharide synthesis enzyme Cap8C	-2.20	16598.52
MW1834	hypothetical protein, similar to ferritin (ftn)	-2.21	42398.52
E161084c SAR1022	V8PROTEASE	-2.22	2995.02
SAV2451	subfamily S9C non-peptidase homologues [putative carboxylesterase]	-2.26	16937.02
SAV0150:Cap5B	capsular polysaccharide synthesis enzyme Cap5B	-2.29	24054.52
MW1754:splB	serine protease	-2.30	1337.52
MW2344:hlgB	gamma-haemolysin component B	-2.31	2088.52
MW1759:bsaE	hypothetical protein, similar to EpiE (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-2.38	8101.52
SAV0157:Cap5I	capsular polysaccharide synthesis enzyme Cap5I	-2.54	10215.52
SAV0160:Cap5L	capsular polysaccharide synthesis enzyme Cap5L	-2.55	10054.52
SAV0164:Cap5P	capsular polysaccharide synthesis enzyme Cap5P	-2.56	12078.52
MW1758:bsaG	hypothetical protein, similar to EpiG (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-2.57	9447.52
SAV0153:Cap8E	capsular polysaccharide synthesis enzyme Cap8E	-2.60	17382.52
MW1757	hypothetical protein, similar to Ear protein [putative exported protein]	-2.61	1243.02
SAV0162:Cap5N	capsular polysaccharide synthesis enzyme Cap5N	-2.62	7946.52
SA1814	subfamily M20A unassigned peptidases [succinyl-diaminopimelate desuccinylase]	-2.64	6931.52
bbp	sdrE homolog (bone sialoprotein binding protein) O24	-2.67	1579.52
SAV0159:Cap5K	capsular polysaccharide synthesis enzyme Cap5K	-2.75	7361.52

SAV0163:Cap8O	capsular polysaccharide synthesis enzyme Cap8O	-2.81	9771.52
SAV0156:Cap5H	capsular polysaccharide synthesis enzyme O-acetyl transferase Cap5H	-2.82	16743.52
MW1753:splC	serine protease	-2.83	1298.02
SA2424	hypothetical protein, similar to transcription regulator Crp/Fnr family protein	-2.86	2042.52
SAR1902	SAR1902 (was splE S.hyicus)	-2.87	1074.52
MW1752:splF	serine protease	-2.89	1282.02
SAV0158:Cap5J	capsular polysaccharide biosynthesis enzyme Cap5J	-2.96	11567.52
SAV1813:splA	subfamily S1B unassigned peptidases (exoprotein A) [serine protease]	-3.01	1804.52
MW2590:lip	triacylglycerol lipase precursor	-3.27	2783.52
MW1032:sdhB	succinate dehydrogenase iron-sulfur protein subunit (MW2)	-3.56	26220.52
MW1910	hypothetical protein [phage protein]	-3.82	2395.52
MW0074	hypothetical protein	-3.89	1488.02
MW1041	hypothetical protein, similar to fibrinogen-binding protein	-3.94	5849.52
efb	fibrinogen-binding protein (strain 4074/MW1040)	-3.96	15162.52
MW2553:arc	carbamate kinase	-3.99	12883.52
MW1040	fibrinogen-binding protein	-4.10	6602.52
MW2431	hypothetical protein, similar to glucarate transporter	-4.13	12962.52
MW0167	hypothetical protein, similar to transcription regulator	-4.54	9809.52
MW0297:geh	glycerol ester hydrolase (lipase)	-5.06	25134.52
MW1881:hlb	beta haemolysin	-5.14	21924.52
MW2343:hlgC	gamma-haemolysin component C	-5.20	5121.02
E161208c SAR1136	BICOMPNTOXIN [hla precursor]	-5.21	8750.52
MW1044:hla	alpha toxin; alpha haemolysin	-5.57	26600.02
MW0396	hypothetical protein [putative exported protein]	-5.81	31226.02
SA0211	acetyl-CoA acetyltransferase probable [thiolase]	-6.29	16199.52

5.1.2.2 Differential expression observed between EE - versus overnight (ON) -growth

Gene expression was also measured after overnight (ON, 21hrs) culture. Changes between the EE and ON cells were investigated. Compared with the 5 v 7.5hr phase, 25 of the 52 genes still had detectable expression levels, but only 7 genes were significantly up-regulated (Table 5.3, marked by astrix, *). These were the putative lantibiotic ABC transporter genes *bsaEFG*, *clfA* (clumping factor A, a fibrinogen binding protein), *sdhB* (an iron related enzyme), MW0396 (a hypothetical, possibly exported protein), and SA0211 (a probable acetyl coA acetyltransferase). With the exception of *clfA* and SA0211, the expression levels of the other 5 genes were greater in the ON phase, indicated by the expression ratios (compare Table 5.2 and 5.3). The increase in expression ratio between the EE to LE and EE to ON phases suggests that these genes are growth phase dependent and that they continue to be up-regulated as cell density increases. An additional set of genes was not found to be up-regulated in the EE to LE comparison but were noted in the EE to ON data (Table 5.3). These included surface binding proteins (e.g. *epbS*, *sdrE*), ribosomal RNA, the protease *clpP*, *dnaK* (a chaperonin) and regulatory proteins including *sarA*. The implied increase in levels of the chaperonin (analogous to hsp70) is likely to be indicative of stress responses in the ON cells that result from nutrient depletion and the accumulation of extracellular toxins. The increase in transcription of the surface proteins was unexpected.

A larger number of genes (n=106) were significantly down-regulated between the EE and ON culture phases (Table 5.4). Here several interesting observations were made. In particular, the expression levels of many of the regulatory genes decreased including the key regulators *agrC1*, *sarH1* (*sarS*) and *sarR*, *sigB* (and its regulatory protein *rsbU*), *saeR* and *arlR*, as well as *lytR*, *vraR*, *ccpA* and several hypothetical proteins with putative regulatory functions. Other down-regulated genes included 15 peptidases, the *hsd*

restriction modification operon (*hsdR/S/M*), two lytic regulatory proteins from transposon Tn554, several transport related proteins and house-keeping genes. In order to determine which of the observed effects were mediated by *agr*, these experiments were repeated using the isogenic mutant strain (SH1000 agr^-).

Table 5.3 Genes up-regulated in the *agr*$^+$ strain between EE and ON growth.

Gene ID	Product description	log_2 ratios (Agr^+ 5vON)	TF
MW1369:epbS	cell surface elastin binding protein	-1.90	11021.07
MW0111	hypothetical protein similar to tetracycline resistance protein [putative transport system protein]	-1.94	1362.07
MWrRNA02	23S ribosomal RNA	-1.97	61362.07
MW0985	spermidine/putrescine-binding protein precursor homolog, potD [ABC transporter]	-2.11	3764.07
SA0573:sarA	staphylococcal accessory regulator A	-2.16	7282.07
SAV0768:clpP	peptidase Clp (type 1) [ATP-dependent Clp protease, proteolytic subunit ClpP]	-2.27	4984.07
SA0857	hypothetical protein, similar to negative regulator of genetic competence MecA [regulatory protein]	-2.31	13315.07
MW1032:sdhB *	succinate dehydrogenase iron-sulfur protein subunit (MW2)	-2.38	5578.07
MWrRNA05	16S ribosomal RNA	-2.87	52414.07
MW1760:bsaF *	hypothetical protein, similar to EpiF (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-3.54	13834.07
MW1758:bsaG *	hypothetical protein, similar to EpiG (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-3.58	5227.07
MW0518:sdrE	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	-3.91	18485.07
MW1759:bsaE *	hypothetical protein, similar to EpiE (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-4.17	5639.07
MW1532:dnaK	DnaK protein [chaperone protein]	-4.86	25581.07
MW0396 *	hypothetical protein [putative exported protein]	-4.99	10119.07
SA2308	hypothetical protein, similar to transcription regulator MarR family	-5.10	3131.07
MW0764:clfA *	fibrinogen-binding protein	-5.20	23461.07
SA0211 *	acetyl-CoA acetyltransferase probable [thiolase]	-5.64	8593.07

Table 5.4 Genes down-regulated in the a*gr*$^+$ strain between EE and ON growth.

Gene ID	Product description	log_2 ratios (Agr^+ 5vON)	TF
MW0723:lgt	prolipoprotein diacylglyceryl transferase	1.93	7240.07
SA1557:ccpA	catabolite control protein A [regulator function also]	1.94	8426.07
SA1649	conserved hypothetical protein	1.98	1399.07
MW1501	family S54 unassigned peptidases (SAV1549 protein aka MW1501) [peptidase, rhomboid family]	2.01	7340.07
MW0339	hypothetical protein, similar to GTP-binding protein [GTP-binding protein YchF (ychF)]	2.02	9592.57
SA1247	truncated (putative two-component response regulator ArlR (truncated-arlR)	2.08	4062.07
SAV1068:purQ	family C56 non-peptidase homologues [phosphoribosylformylglycinamidine synthase]	2.10	1321.07
MW0573	hypothetical protein similar to iron-binding protein [iron compound ABC transporter, permease protein/ lipoprotein]	2.12	1432.07
SAV1384	oligopeptidase F (pepF)	2.14	1811.07
SAV1057:fmt	family S12 unassigned peptidases (Fmt protein) [autolysis and methicillin-resistant-related protein (fmt)]	2.14	1026.07
SA0675	hypothetical protein, similar to ABC transporter ATP-binding protein	2.14	1769.07
SA1320	hypothetical lipoprotein	2.17	1513.07
SA1248	truncated (putative response regulator ArlR [S (truncated-arlR)]	2.18	2662.57
SA1748	hypothetical protein, similar to transcription regulator, GntR family	2.20	3372.07
MW1163	hypothetical protein, similar to 3-oxoacyl- acyl-carrier protein reductase homolog ymfI	2.22	2408.07
SA0661:saeR	response regulator	2.24	5931.07
MW0213	hypothetical protein, similar to nickel ABC transporter nickel-binding protein [putative extracellular solute-binding lipoprotein]	2.25	1083.57
SA1329	ferric uptake regulator (fur) homolog [transcriptional regulator, Fur family]	2.26	946.073
MW1594:obg	Spo0B-associated GTP-binding protein	2.27	4482.07
SA1700:vraR	two-component response regulator	2.30	1902.57
MW2403	hypothetical protein, similar to oxidoreductase	2.33	2493.07
MW1872	hypothetical protein, similar to ABC transporter, ATP-binding protein	2.34	2643.07
MW0795	ABC transporter ATP-binding protein homologue	2.34	8330.07
MW0736:tpi	triosephosphate isomerase	2.36	10517.07
MW1874	hypothetical protein, similar to ABC transporter, ATP-binding protein	2.42	1835.07
SA2272	hypothetical protein	2.47	1087.57

SAV1279	subfamily M16C unassigned peptidases [putative protease]	2.48	3158.57
SA1666	two-component response regulator homolog [DNA-binding response regulator, LuxR family]	2.49	908.073
SAV2133:hmrA	subfamily M20D non-peptidase homologues (HmrA protein) [puative peptidase/ similar to amidase (HmrA)]	2.50	2002.07
SA0702:llm	lipophilic protein affecting bacterial lysis rate and methicillin resistance level [lipophilic regulator protein]	2.51	1209.07
SA2418	hypothetical protein, similar to two-component response regulator	2.53	998.073
MW0623:vraF	ABC transporter ATP-binding protein	2.54	1268.07
MW1572	iron-sulfur cofactor synthesis protein homolog	2.56	3248.07
MW1536:lepA	GTP-binding protein	2.62	1782.07
MW0558	conserved hypothetical protein [putative membrane protein]	2.64	971.073
MW1547:aroE	shikimate dehydrogenease	2.64	2591.07
SA0298	hypothetical protein, similar to regulatory protein PfoR (perfringolysin O regulator protein (pfoR))	2.68	2027.07
SAV0432:hsdS	probable restriction modification system specificity subunit c	2.77	1410.07
MW2103	hypothetical protein similar to ferrichrome ABC transporter (binding protein)	2.80	2407.07
MW0077	hypothetical protein, similar to transcription regulator AraC/XylS family	2.87	1173.07
SA0726:gapR	glycolytic operon regulator [transcriptional regulator]	2.90	13755.07
MW2013	lipoprotein precursor [membrane protein oxaA precursor (oxaA)]	2.90	3469.07
MW1988:sigB	sigma factor B	2.90	15212.07
SAV0511:ftsH	FtsH-2 peptidase [cell-division protein (ftsH)]	2.91	10016.07
MW2328	hypothetical protein similar to Zn-binding lipoprotein adcA [ABC transporter, substrate-binding protein/ probable zinc-binding lipoprotein]	2.99	991.573
SA2421	hypothetical protein, similar to transcriptional regulator	3.03	4273.07
SA2340	hypothetical protein, similar to transcriptional regulator tetR-family	3.08	1777.57
MW1828:map	methionyl aminopeptidase map	3.08	3748.07
SA0276	conserved hypothetical protein, similar to diarrheal toxin yukA [essC protein]	3.09	906.07
SAV1044	family S33 unassigned peptidases [hydrolase, alpha/beta hydrolase fold family]	3.13	4111.57
MW2197:modA	probable molybdate-binding protein	3.13	4296.07
SA1676	hypothetical protein, similar to regulatory protein (pfoS/R)	3.15	987.07
gi|78172212 (agr)	agr type 1 subset specific 1 12 16 17 20 25	3.16	11804.07
MW0108	hypothetical protein [cell wall surface anchor family protein]	3.17	1957.07
MW1971:vga	hypothetical ABC transporter ATP-binding protein	3.20	1040.07
MW2409	conserved hypothetical protein [putative helicase]	3.22	997.07

SAV1113	conserved hypothetical protein [cell division protein, FtsW/RodA/SpoVE family]	3.23	1195.57
SA1956	lytic regulatory protein truncated with Tn554	3.24	2334.07
SA1139:glgP	glycerol uptake operon antiterminator regulatory protein	3.29	2236.07
SA1949	lytic regulatory protein truncated with Tn554 (truncated-SA) [membrane-embedded lytic regulatory protein]	3.31	4731.07
MW1064:pbpA	penicillin-binding protein 1	3.35	2831.07
MW0543:pta	phosphotransacetylase	3.36	8621.07
MW1092:gmk	guanylate kinase	3.38	12260.07
MW1942	hypothetical protein similar to leukocidin chain lukM precursor [probable leukocidin S subunit/ Aerolysin/Leukocidin family protein]	3.39	1059.57
MW0682	hypothetical protein, similar to ABC transporter ATP-binding protein	3.46	959.07
MW2259	conserved hypothetical protein [YnfA family protein/ putative membrane protein]	3.48	1683.07
SAV0965:spsB	signal peptidase SpsB (SpsB protein)	3.50	1994.07
MW2336	hypothetical protein, similar to amino acid ABC transporter, periplasmic amino acid-binding protein	3.56	9466.07
SAV1262	RseP peptidase	3.57	3165.57
SAV0195:hsdR	probable type I restriction enzyme restriction chain	3.57	1066.07
MW2368	hypothetical protein, similar to integral membrane efflux protein [putative drug transporter]	3.59	1406.07
SA0704	conserved hypothetical protein [degV family protein]	3.61	1734.07
agrC1	group 1 accessory gene regulator, agr operon (af210055)	3.71	11071.07
MW0263	conserved hypothetical protein, similar to diarrheal toxin incomplete ORF	3.71	1270.07
MW0698	lipoprotein, similar to ferrichrome ABC transporter [also transferrin receptor]	3.71	3349.07
MW2261	ATP-binding protein) hypothetical protein, similar to ABC transporter	3.71	1771.07
MW0342:ssb	single-strand DNA-binding protein of phage phi Sa 2mw	3.74	16296.07
MW1362:hu	DNA-binding protein II	3.7	17837.07
SAV2051	family M22 non-peptidase homologues [similar to glycoprotein endopeptidase]	3.94	1029.07
SA0509:HCHA	family C56 unassigned peptidases [DJ-1/PfpI family protein]	3.95	12933.07
SA0879:htrA	subfamily S1C unassigned peptidases [serine protease (htrA)]	3.95	1491.07
MW0085:sarH1	staphylococcal accessory regulator A homologue	4.00	1939.07
MW0552	major tail protein [putative membrane protein NOT major tail protein]	4.00	1679.57
MW0574	hypothetical protein similar to iron(III) ABC transporter permease protein	4.05	1072.07
MW1991:rsbU	sigmaB regulation protein RsbU	4.08	13227.07
MW0323	conserved hypothetical protein [putative Sec-independent protein translocase protein]	4.09	922.07
MW2195:modC	molybdenum transport ATP-binding protein [molybdenum ABC transporter, ATP-binding protein ModC]	4.13	3632.07
MW0919	hypothetical protein, similar to UDP-glucose:polyglycerol phosphate glucosyltransferase	4.15	2590.07
MW2102	hypothetical protein similar to ferrichrome ABC transporter (permease)	4.15	998.07

SAV0431:hsdM	probable type I site-specific deoxyribonuclease LldI chain	4.17	2634.57
SA0641	conserved hypothetical protein, similar to transcriptional regulator [probable transcriptional regulator MarR family]	4.25	10974.07
SA2144	hypothetical protein, similar to transcriptional regulator (TetR/AcrR family)	4.29	2767.07
MW1280	ABC transporter (ATP-binding protein) homolog	4.30	1840.07
SAV1529	subfamily M24B non-peptidase homologues [proline dipeptidase]	4.34	8225.07
SAV1751	subfamily M20A unassigned peptidases	4.38	14684.07
MW1482	subfamily M24B unassigned peptidases [proline dipeptidase]	4.39	7299.07
MW0595	hypothetical protein, similar to ABC transporter ATP-binding protein	4.45	4460.07
MW2273	hypothetical protein, similar to multidrug resistance protein [drug resistance transporter, EmrB/QacA subfamily, putative]	4.48	1203.07
MW1811	hypothetical protein, similar to teichoic acid translocation ATP-binding protein tagH [ABC transporter]	4.70	966.07
SA1360	subfamily M24B unassigned peptidases [proline dipeptidase]	4.75	5982.07
MW1364	hypothetical protein, similar to GTP binding protein [putative GTPase/ GTP-binding protein EngA (engA)]	4.78	5116.07
MW0593	lipoprotein, Streptococcal adhesin PsaA homologue	4.81	2783.07
SAV1000	oligopeptidase F (pepF)	5.24	2688.57
SAV1879:ampS	aminopeptidase S (aminopeptidase ampS)	5.40	4423.07
MW1182:glpF	glycerol uptake facilitator	5.45	5116.07
SA0251:lytR	two-component response regulator [sensory transduction protein LytR]	5.87	2869.07
MW2213:sarR	staphylococcal accessory regulator A homolog	5.91	11903.57
MW0594	conserved hypothetical protein [ABC transporter permease protein]	6.14	2305.07
MW0084:spa	IgG binding protein A	6.47	15225.07

5.1.3 *Agr*⁻ (mutant) strain

The *agr* mutant strain was grown and sampled at the same time points. Overall, less expression at all time phases and comparisons was observed.

5.1.3.1 Differential expression observed between EE- versus LE-growth

Only six genes were found to be significantly up-regulated (Table 5.5) between the EE and LE cultures. As observed in the *agr*$^{+}$ strain, increased expression of *hla* (MW1044), *hla* precursor (SAR1136) and MW0396 (a putative exported protein) were observed in the LE phase. As these genes are up-regulated in the *agr* mutant strain, the up-regulation was independent of the *agr* operon. Two other genes, *pyrC* (dihydro-orotase) and *arcC* (carbamate kinase, the housekeeping gene used in MLST) were also up-regulated between these phases. As for *hla*, the regulation was independent of the presence of RNAIII.

Table 5.5 Genes up-regulated in the *agr*⁻ strain between EE and LE growth.

Gene ID	Product description	$\log_2$ ratios (Agr⁻ 5v7.5)	TF
MW2553:arc	carbamate kinase [arcC]	-2.22	3125.53
MW0396	hypothetical protein [putative exported protein]	-2.40	17318.53
MW1044:hla	alpha toxin; alpha haemolysin	-2.40	10994.53
SAV1201:pyrC	dihydro-orotase	-2.52	6595.533
E161208c SAR1136	bicompntoxin [hla precursor]	-2.95	1673.533
SA0211	acetyl-CoA acetyltransferase probable [thiolase]	-5.43	17726.53

Similarly, comparative data show another six genes were significantly down-regulated between the EE and LE phases (Table 5.6). Four of these genes were hypothetical proteins, which showed similarity to (*i*) an ABC transporter (MW2336), (*ii*) lysR family transcriptional regulator (SA2123), (*iii*) maltose-binding protein (MW0190) and (*iv*) a

multi-drug resistance protein (MW2273). The other two genes were a peptidase (SA0509) and a myosin-cross reactive streptococcal antigen homologue (SA0102). Significant expression of these genes was not detected in the *agr*$^+$ strain although the ABC transporter (MW2336) and the regulator gene (SA2123) were detected at low levels. Therefore, it was not possible to determine whether expression of these genes was growth phase dependent in the *agr*$^+$ strain. However, the results are evidence that certain genes are repressed by the Agr operon in EE cells. It may be that this class of genes are very sensitive to RNAIII.

Table 5. 6 Genes down-regulated in the *agr*$^-$ strain between EE and LE growth.

Gene ID	Product description	log_2 ratios (Agr$^-$ 5v7.5)	TF
SA0509: HCHA	family C56 unassigned peptidases	3.26	27860.53
MW2273	hypothetical protein, similar to multidrug resistance protein	2.13	1236.53
SA2123	hypothetical protein, similar to transcription regulator LysR family	2.24	2499.53
SA0102	67 kDa Myosin-crossreactive streptococcal antigen homologue	2.46	929.53
MW2336	hypothetical protein, similar to ABC transporter, periplasmic amino acid-binding protein	3.21	11224.53
MW0190	hypothetical protein, similar to maltose/maltodextrin-binding protein	3.30	1304.53

5.1.3.2 Differential expression observed between the EE- versus ON-growth

Comparison of the EE and ON phase RNAs revealed nine genes that were significantly up-regulated (Table 5.7); *clfA* (MW0264), a probable acetyl-coA acetyl transferase (SA0211), *bsaE* (MW1759) and *bsaG* (MW1758; putative lantibiotic epidermin immunity proteins/ ABC transporters), a peptidase, exotoxin 18 (*set18*), peptidase *SAV1512*, intracellular adhesin protein C (*icaC*) and serine proteases *splA* (MW1755) and *splB* (MW1754). Those

marked by an astrix also displayed up-regulation in the isogenic *agr*$^+$ strain between the same time points (Table 5.3).

Table 5.7 Genes up-regulated in the *agr*$^-$ strain between EE and ON growth.

Gene ID	Product description	log_2 ratios (Agr- 5vON)	TF
SAV1512	subfamily M20B unassigned peptidases	-1.98	3762.58
SAV2669:icaC	intercellular adhesion protein C	-2.14	1062.08
MWrRNA05 *	16S ribosomal RNA (RNA05)	-2.22	80665.08
MW1758:bsaG *	hypothetical protein, similar to EpiG (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-2.57	4063.58
MW1754:splB	serine protease	-2.86	994.08
MW1759:bsaE *	hypothetical protein, similar to EpiE (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-2.88	3904.08
MW0384:set18	exotoxin 18	-3.49	1625.08
MW0764:clfA *	fibrinogen-binding protein	-3.95	25515.08
MW1532:dnaK *	DnaK protein	-4.33	21299.08
MW1755:splA	serine protease	-4.48	868.58
SA0211 *	acetyl-CoA acetyltransferase probable [thiolase]	-6.75	10067.08

As found for the *agr*$^+$ strain, a larger number of genes were down-regulated between EE and ON growth phases. Here 54 genes were noted (Table 5.8), with the highest ratios observed for *spaA*, *glpF* (glycerol uptake facilitator) and the regulator *sarR*.

Table 5.8 Genes down-regulated in the *agr*⁻ strain between EE and ON growth.

Gene ID	Product description	log2 ratios (Agr-5vON)	TF
MW2195:modC	molybdenum transport ATP-binding protein	2.054331173	3163.08
MW1163	hypothetical protein, similar to 3-oxoacyl- acyl-carrier protein reductase homolog ymfI	2.067582897	2462.58
SA2340	hypothetical protein, similar to transcriptional regulator tetR-family	2.081029416	1965.08
MW1594:obg	Spo0B-associated GTP-binding protein	2.089941575	5809.08
SAV0511:ftsH	FtsH-2 peptidase	2.123756399	7353.58
SAV1044	family S33 unassigned peptidases	2.136806483	5115.08
MW2197:modA	probable molybdate-binding protein	2.144910727	3081.08
MW0790	ABC transporter ATP-binding protein homologue	2.161505009	1351.08
MW0200	hypothetical protein similar to periplasmic-iron-binding protein BitC	2.183270352	1191.08
SAV1384	oligopeptidase F	2.186211334	2491.08
SAV2051	family M22 non-peptidase homologues	2.209133559	1029.08
MW0339	hypothetical protein, similar to GTP-binding protein	2.256324463	9048.08
MW1988:sigB	sigma factor B	2.267941227	16231.08
MW0871:OppF	oligopeptide transport ATP-binding protein	2.299083297	4382.08
MW1092:gmk	guanylate kinase	2.305221603	14291.58
SAV1279	subfamily M16C unassigned peptidases	2.312880849	2287.08
SA1248	truncated (putative response regulator ArlR [S (truncated-arlR)]	2.376545244	3884.08
SA1247	truncated (putative response regulator ArlR (truncated-arlR)]	2.380709953	7388.08
MW1572	iron-sulfur cofactor synthesis protein homolog	2.384973489	5390.08
SA0879:htrA	subfamily S1C unassigned peptidases	2.39007265	1269.08
SAV0431:hsdM	probable type I site-specific deoxyribonuclease LldI chain	2.404862071	2438.08
SA0276	conserved hypothetical protein, similar to diarrheal toxin	2.418257877	942.08
SAV1068:purQ	family C56 non-peptidase homologues (phosphoribosylformylglycinamidine synthase)	2.457022941	1710.08
SAV0195:hsdR	probable type I restriction enzyme restriction chain	2.465547286	1559.08
SAV1398:hipO	family M20D unassigned peptidases (hippurate hydrolase)	2.556010903	1709.08
SA1557:ccpA	catabolite control protein A (regulator function also)	2.597707603	11970.08
SA1700:vraR	two-component response regulator	2.616065258	1644.08
MW0108	hypothetical protein	2.655320538	8422.08
MW2261	ATP-binding protein) hypothetical protein, similar to ABC transporter	2.714697528	2804.08
MW0342:ssb	single-strand DNA-binding protein of phage phi Sa 2mw (phage protein)	2.744794791	8317.08

SAV1262	RseP peptidase	2.769100533	1628.58
MW0543:pta	phosphotransacetylase	2.837414254	9089.08
SA0641	conserved hypothetical protein, similar to transcriptional regulator	2.840251747	5921.08
MW0795	ABC transporter ATP-binding protein homologue	2.860950931	10628.08
MW1044:hla	alpha toxin; alpha haemolysin	2.88377907	1203.08
MW0870:OppD	oligopeptide transport system ATP-binding protein OppD homologue	2.90446186	1210.58
MW1362:hu	DNA-binding protein II	2.961804985	27765.08
MW2336	hypothetical protein, similar to ABC transporter, periplasmic amino acid-binding protein	2.971148981	9957.08
SAV0520	family C56 non-peptidase homologues	3.005459543	1967.08
MW0919	hypothetical protein, similar to UDP-glucose:polyglycerol phosphate glucosyltransferase	3.040339866	2731.08
SA1360	subfamily M24B unassigned peptidases	3.076465764	4785.08
MW1482	subfamily M24B unassigned peptidases	3.185571518	7706.08
SAV1000	oligopeptidase F	3.189526652	3002.58
MW2431	hypothetical protein, similar to glucarate transporter	3.19224756	17977.08
SAV1529	subfamily M24B non-peptidase homologues	3.24835505	7075.08
MW1834	hypothetical protein, similar to ferritin	3.305430881	19237.08
MW1991:rsbU	sigmaB regulation protein RsbU	3.366225659	18825.08
SA0661:saeR	response regulator	3.502727232	2496.58
SAV1751	subfamily M20A unassigned peptidases	3.678418791	19358.08
SA0509:HCHA	family C56 unassigned peptidases	3.941352888	18603.08
MW0084:spa	IgG binding protein A	4.383021426	40797.58
MW1182:glpF	glycerol uptake facilitator	4.924044687	31560.08
MW2213:sarR	staphylococcal accessory regulator A homolog	5.07577239	18945.08

5.1.4 Agr summary

Once all significant measurements were recorded, the data from the isogenic strains were compared to determine those genes with possible *agr*-related effects. The *agr* gene product is known to be centrally involved in quorum sensing, that is, responding to the environment in a growth-phase dependent manner. Simplistically, any gene regulated via *agr* would not be expected to show the same effect (either no regulation, or reduced regulation) in the mutant strain. However, due to the complexity of the regulatory network in *S. aureus*, further work would be necessary to confirm *agr* involvement, and the mechanisms involved (e.g. direct/ indirect *agr* effect or possibly involving other regulators).

In the first comparison (Table 5.3 versus Table 5.7), the genes upregulated between EE and ON growth phases in the isogenic strains were compared. Those genes that appeared in both lists were excluded (the presence of these genes in the agr^- mutant strain essentially eliminates the possibility of *agr* regulation), and the raw data for those remaining in the agr^+ strain were re-analysed. The revised list is shown in Table 5.9. The highlighted genes appear to be strongly associated with the presence of *agr* (that is, the isogenic strains show opposite effects), whilst the others have a weak association (the presence of *agr* enhances the overall effect).

In the same way, the isogenic strains were also compared for those genes which could be negatively regulated by *agr* (compare Table 5.4 and Table 5.8). Following the same algorithm as above, the results obtained are summarised in Table 5.10.

Table 5.9 Genes positively regulated by *agr* according to the data derived from this study.

The negative log ratios indicate higher expression in the ON phase; in this case this gives a measure of gene expression in response to quorum sensing. The ratios of -3 and below suggest stronger responses to quorum sensing.

Gene ID	Product description	log_2 ratios (Agr^+ 5vON)	TF
MW1369:epbS	cell surface elastin binding protein	-1.90	11021.07
MW0111	hypothetical protein similar to tetracycline resistance protein [putative transport system protein]	-1.94	1362.07
MW0985	spermidine/putrescine-binding protein precursor homolog, potD [ABC transporter]	-2.11	3764.07
SA0573:sarA	staphylococcal accessory regulator A	-2.16	7282.07
SAV0768:clpP	peptidase Clp (type 1) [ATP-dependent Clp protease, proteolytic subunit ClpP]	-2.27	4984.07
SA0857	hypothetical protein, similar to negative regulator of genetic competence MecA [regulatory protein]	-2.31	13315.07
MW1032:sdhB	succinate dehydrogenase iron-sulfur protein subunit (MW2)	-2.38	5578.07
MW1760:bsaF	hypothetical protein, similar to EpiF (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	-3.54	13834.07
MW0518:sdrE	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	-3.91	18485.07
MW0396	hypothetical protein [putative exported protein]	-4.99	10119.07
SA2308	hypothetical protein, similar to transcription regulator MarR family	-5.10	3131.07

Table 5.10 Genes negatively regulated by *agr* according to the data generated from this study.

Gene ID	Product description	log_2 ratios (Agr^+ 5vON)	TF
MW0723:lgt	prolipoprotein diacylglyceryl transferase	1.93	7240.07
SA1649	conserved hypothetical protein	1.98	1399.07
MW1501	family S54 unassigned peptidases (SAV1549 protein aka MW1501) [peptidase, rhomboid family]	2.01	7340.07
MW0573	hypothetical protein similar to iron-binding protein [iron compound ABC transporter, permease protein/ lipoprotein]	2.12	1432.07
SAV1057:fmt	family S12 unassigned peptidases (Fmt protein) [autolysis and methicillin-resistant-related protein (fmt)]	2.14	1026.07
SA0675	hypothetical protein, similar to ABC transporter ATP-binding protein	2.14	1769.07
SA1320	hypothetical lipoprotein	2.17	1513.07
SA1748	hypothetical protein, similar to transcription regulator, GntR family	2.20	3372.07
MW0213	hypothetical protein, similar to nickel ABC transporter nickel-binding protein [putative extracellular solute-binding lipoprotein]	2.25	1083.57
SA1329	ferric uptake regulator (fur) homolog [transcriptional regulator, Fur family]	2.26	946.07
MW2403	hypothetical protein, similar to oxidoreductase	2.33	2493.07
MW1872	hypothetical protein, similar to ABC transporter, ATP-binding protein	2.34	2643.07
MW0736:tpi	triosephosphate isomerase	2.36	10517.07
MW1874	hypothetical protein, similar to ABC transporter, ATP-binding protein	2.42	1835.07
SA2272	hypothetical protein	2.47	1087.57
SA1666	two-component response regulator homolog [DNA-binding response regulator, LuxR family]	2.49	908.07
SAV2133:hmrA	subfamily M20D non-peptidase homologues (HmrA protein) [puative peptidase/ similar to amidase (HmrA)]	2.50	2002.07
SA0702:llm	lipophilic protein affecting bacterial lysis rate and methicillin resistance level [lipophilic regulator protein]	2.51	1209.07
SA2418	hypothetical protein, similar to two-component response regulator	2.53	998.07
MW0623:vraF	ABC transporter ATP-binding protein	2.54	1268.07
MW1536:lepA	GTP-binding protein	2.62	1782.07

MW0558	conserved hypothetical protein [putative membrane protein]	2.64	971.07
MW1547:aroE	shikimate dehydrogenease	2.64	2591.07
SA0298	hypothetical protein, similar to regulatory protein PfoR (perfringolysin O regulator protein (pfoR))	2.68	2027.07
SAV0432:hsdS	probable restriction modification system specificity subunit c	2.77	1410.07
MW2103	hypothetical protein similar to ferrichrome ABC transporter (binding protein)	2.81	2407.07
MW0077	hypothetical protein, similar to transcription regulator AraC/XylS family	2.87	1173.07
SA0726:gapR	glycolytic operon regulator [transcriptional regulator]	2.90	13755.07
MW2013	lipoprotein precursor [membrane protein oxaA precursor (oxaA)]	2.90	3469.07
MW2328	hypothetical protein similar to Zn-binding lipoprotein adcA [ABC transporter, substrate-binding protein/ probable zinc-binding lipoprotein]	2.99	991.57
SA2421	hypothetical protein, similar to transcriptional regulator	3.03	4273.07
MW1828:map	methionyl aminopeptidase map	3.09	3748.07
SA1676	hypothetical protein, similar to regulatory protein (pfoS/R)	3.15	987.07
gi|78172212 (agr)	agr type 1 subset specific 1 12 16 17 20 25	3.16	11804.07
MW1971:vga	hypothetical ABC transporter ATP-binding protein	3.20	1040.07
MW2409	conserved hypothetical protein [putative helicase]	3.22	997.07
SAV1113	conserved hypothetical protein [cell division protein, FtsW/RodA/SpoVE family]	3.23	1195.57
SA1956	lytic regulatory protein truncated with Tn554	3.25	2334.07
SA1139:glgP	glycerol uptake operon antiterminator regulatory protein	3.29	2236.07
SA1949	lytic regulatory protein truncated with Tn554 (truncated-SA) [membrane-embedded lytic regulatory protein]	3.31	4731.07
MW1064:pbpA	penicillin-binding protein 1	3.35	2831.07
MW1942	hypothetical protein similar to leukocidin chain lukM precursor [probable leukocidin S subunit/ Aerolysin/Leukocidin family protein]	3.39	1059.57
MW0682	hypothetical protein, similar to ABC transporter ATP-binding protein	3.46	959.07
MW2259	conserved hypothetical protein [YnfA family protein/ putative membrane protein]	3.48	1683.07
SAV0965:spsB	signal peptidase SpsB (SpsB protein)	3.50	1994.07
MW2368	hypothetical protein, similar to integral membrane efflux protein [putative drug transporter]	3.60	1406.07
SA0704	conserved hypothetical protein [degV family protein]	3.61	1734.07
agrC1	group 1 accessory gene regulator, agr operon (af210055)	3.71	11071.07
MW0263	conserved hypothetical protein, similar to diarrheal toxin incomplete ORF	3.71	1270.07
MW0698	lipoprotein, similar to ferrichrome ABC transporter [also transferrin receptor]	3.71	3349.07

MW0085:sarH1	staphylococcal accessory regulator A homologue	4.00	1939.07
MW0552	major tail protein [putative membrane protein NOT major tail protein]	4.00	1679.57
MW0574	hypothetical protein similar to iron(III) ABC transporter permease protein	4.05	1072.07
MW0323	conserved hypothetical protein [putative Sec-independent protein translocase protein]	4.09	922.07
MW2102	hypothetical protein similar to ferrichrome ABC transporter (permease)	4.15	998.07
SA2144	hypothetical protein, similar to transcriptional regulator (TetR/AcrR family)	4.29	2767.07
MW1280	ABC transporter (ATP-binding protein) homolog	4.30	1840.07
MW0595	hypothetical protein, similar to ABC transporter ATP-binding protein	4.45	4460.07
MW2273	hypothetical protein, similar to multidrug resistance protein [drug resistance transporter, EmrB/QacA subfamily, putative]	4.48	1203.07
MW1811	hypothetical protein, similar to teichoic acid translocation ATP-binding protein tagH [ABC transporter]	4.70	966.07
MW1364	hypothetical protein, similar to GTP binding protein [putative GTPase/ GTP-binding protein EngA (engA)]	4.78	5116.07
MW0593	lipoprotein, Streptococcal adhesin PsaA homologue	4.81	2783.07
SAV1879:ampS	aminopeptidase S (aminopeptidase ampS)	5.40	4423.07
SA0251:lytR	two-component response regulator [sensory transduction protein LytR]	5.87	2869.07
MW0594	conserved hypothetical protein [ABC transporter permease protein]	6.15	2305.07

Although these results demonstrate the anticipated involvement of the *agr* operon in gene regulation, the mechanism involving RNAIII-related quorum sensing (usually triggered during cellular growth) was not particularly evident in the test strains. The observed changes in expression were apparently unrelated to the intracellular concentration of RNAIII since this transcript (although present in a relatively high levels) was not noted in any of the up- or down-regulated gene lists (see above). Nevertheless, RNAIII involvement cannot be ruled out.

This apparent non-response of RNAIII to quorum sensing may be an adapted feature of this common laboratory strain. NCTC 8325-4 has been used in laboratory research for many years, and presumably has accumulated genetic changes during repeated sub-culture. Some of these changes should confer an advantage status to this strain in this environment (e.g. adaptations that favour *in vitro* growth). One such example may be the constant expression of RNAIII such that factors involved in cellular proliferation and attachment (controlled via RNAIII) are continuously expressed instead of following traditional trends of expression. If this was the case, it would seem that the re-introduction of *rsbU* into strain NCTC 8325 (SH1000) has not corrected this lack of RNAIII-mediated quorum sensing. Sabersheikh and Saunders (2004) in studies of the virulence-associated gene transcripts of *S. aureus* also noted lack of a clear pattern of RNAIII transcription. Fluctuations in RNAIII levels were observed, and thought to be correlated with irregular growth patterns. This same irregular growth was also observed for strains SH1000 and SH1001 in this study (section 3.1.2.1.2). Sabersheikh and Saunders (2004) proposed that the mechanism by which RNAIII is induced is more complex than originally anticipated and that variations in this mechanism exist between different *S. aureus* clonotypes. It is therefore possible that contiguous expression of RNAIII is a feature of this genetic background. Another possibility is that a technical limitation prevented accurate quantification of RNAIII i.e. the high RNAIII levels may have been beyond the limit of

quantification of the array so that any changes in gene expression were not detected. However, pixel saturation was not observed for RNAIII, while this did occur for the rRNA probes suggesting that the results were reliable.

5.2 Biofilm study

5.2.1 Background

Staphylococcus aureus and *S. epidermidis* present a significant clinical problem in device-related infections (O'Gara and Humphreys, 2001). *S. aureus* infections associated with implanted microbial devices or foreign bodies, as well as endocarditis and osteomylitis are associated with biofilm formation. Biofilms can be defined as a "*community of microorganisms often adhered to a surface and encased in an extracellular polysaccharide matrix or glycocalyx*" (Kumamoto and Vinces 2005, O'Toole *et al.*, 2000). Given that hospital inpatients may have foreign devices *in situ* (e.g. catheters and artificial heart valves), colonization and infection with staphylococci (often multi-resistant) can be problematic. The challenge associated with these microbial communities is their inefficient response to control with standard antibiotics (Ceri *et al.*, 1999) and their resistance to targeted host defence mechanisms (Donlan *et al.* 2002, Leid *et al.* 2002).

Recently, it has been shown that these microbial communities often include subpopulations harbouring a diverse set of phenotypes (termed 'variants'). The characteristics of these subpopulations may have overall beneficial effects on the biofilm community (Yarwood *et al.*2007, Boles *et al.* 2004). Associated with this finding is the key hypothesis that the presence of a diverse population within a biofilm will enable the overall community to survive a much broader range of conditions than would a less complex community displaying fewer phenotypes (Boles *et al.* 2004).

Biofilm formation is a multistage process dependent on several factors including the type of bacterial species present, the surface composition (i.e. the material to which attachment occurs), environmental stimuli and the expression of genes essential to the process

(Carpentier and Cerf 1993, Dunne 2002). The primary stage involves microbial adhesion to a surface, which has been recognised to occur via specific molecular mechanisms. Upon insertion of a foreign medical device into a patient, the material quickly become coated with a conditioning film of primarily host-derived extracellular matrix proteins. It is these proteins that may act as receptors for microbial attachment. Staphylococcal attachment to extracellular matrix proteins is facilitated by microbial surface components recognising adhesive matrix molecules (MSRAMMs) (Foster and Hook, 1998). Following attachment, cellular communication occurs resulting in biofilm formation (O'Gara, 2007). Bacteria attached to a surface aggregate to form colonies and become encapsulated in an exo-polysaccharide matrix, termed a gylcocalyx (Costerton *et al.*, 1999).

Cell-to-cell communication (quorum sensing), a critical component of biofilm formation, is the process by which cells communicate with their neighbours (Jones, 2005). Cellular communication within a biofilm community is essential for the coordinated behaviour in response to environmental stimuli or stress. A key quorum sensing mechanism in *S. aureus* is mediated by the accessory gene regulator. The Agr system is a key regulator in *S. aureus* pathogenesis, and exerts control over several genes important for biofilm development such as surface-associated adhesins e.g. fibronectin binding protein. With increasing cell density, *agr* represses the expression of surface-associated adhesins (that mediate cell attachment) and increases the expression of secreted proteins that mediate detachment (e.g. delta toxin, a molecule with surfactant-like properties) and tissue-degrading enzymes (e.g. proteases and haemolysins) to favour bacterial spread. It has been observed that *agr* mutants form more robust biofilms (*in vivo* and *in vitro*) compared with their wild type counterparts (Vuong *et al.* 2000, Vuong *et al.* 2004, and Yarwood *et al.* 2004).

The *ica* operon-encoded enzymes are the best understood factors implicated in biofilm formation in staphylococci (Mack *et al.* 1996, Maria-Litran *et al.* 2002). The mechanism involves production of an extracellular polysaccharide adhesin known as polysaccharide intracellular adhesin (PIA) or polymeric N-acetyl-glucosamine (PNAG). Control of PIA production is mediated by the genes of the *ica*ADBC operon (Heilmann *et al.* 1996, Yazdani *et al.* 2006). Although strain variation has been documented (Cramton *et al.* 1999, Rohde *et al.* 2001, Beenken *et al.* 2004), the majority of *S. aureus* strains contain the *ica* operon (Cramton *et al.* 1999, Fowler *et al.* 2001, Rohde *et al.* 2001). The product of *icaA* is a transmembrane protein with homology to N-acetyl-glucosaminyltransferase. Transcription of *icaA* is controlled by the IcaD protein (Gerke *et al.* 1998). The *icaAD* produced N-acetyl-glucosamine oligomers (polysaccharide chain) reach a maximal length of 20 residues and are only elongated further when *icaAD* is co-expressed with *icaC*, a putative transmembrane protein (Gerke *et al.* 1998). It is likely that translocation of the growing polysaccharide to the cell surface involves the *icaC* gene product. Finally, a surface attached protein (IcaB) mediates deacetylation of the poly-N-acetyl-glucosamine molecule (Vuong *et al.* 2004) permitting bacterial attachment to the cell surface and thereby promoting biofilm development (Vuong *et al.* 2004, O'Gara *et al.* 2007). Regulation of *ica* expression and biofilm formation involves multiple regulatory elements including *agr* and *sarA* (Beenken *et al.*, 2004). In particular *SarA* mutants have a greatly reduced ability to form biofilms (Beenken *et al.*, 2004).

In the past, research into microbial growth primarily focused on the study of monotypic 'free growing' or planktonic organisms in liquid media. However, more recently these studies have moved towards the analysis of microbial growth in biofilms, that is, multi-layer cellular communities (Percival and Bowler 2004, Wilson 2001). Such research has highlighted the marked differences between cellular behaviour, structure and physiology in the two states.

5.2.2. Study results

Biofilm development is thought to be controlled via gene expression that differs significantly between biofilm-associated and planktonic organisms (Beloin and Ghigo, 2005). In this study, *S. aureus* cells grown in a continuous culture biofilm were analysed in parallel with the same strain grown in suspension. The planktonic cells were isolated in the early exponential and stationary phases for comparative analysis. As described in the methods (section 2.2.1.2), biofilm growth was supported over a seven day period in the CDFF instrument, whilst the planktonic cells were grown continuously for 21 hours, with sampling at 5 hours and 21 hours. The data represent RNA transcripts isolated at the time of sampling. The biofilms were grown over a period of 7 days, however, it was not allowed to expand due to the continuous removal of material protruding above the plane of the device and the nutrient medium was maintained. Analysis of the data shows genes expressed differentially between the two states. A high cut-off threshold was set such that only genes with expression levels above this threshold were analysed. Although this leads to the elimination of weak positive results a high threshold was necessary to ensure that the results were robust.

In the analysis of virulence-associated genes, the biofilm state represents a model of infection in which attached, densely-packed cells, communicate for collective survival against host factors. It is a simplistic model for biofilm growth, reflecting biofilm-associated genes involved in the colonisation process and cell-to-cell communication. The dense packing of cells also means neighbouring cells are affected by each others exo-enzymes. However, in this particular model, the biofilm layer was maintained at a constant thickness. Nutrient availability was maintained through continuous supply, more closely reflecting the situation *in vivo* than batch culture.

Planktonic free-living cells are less likely to be affected by the by-products of neighbouring cells, at least during the early phase of culture. During early growth, planktonic cells grow rapidly due to high nutrient availability. However, during latter stages, cell growth rates stall presumably due to restricted nutrient levels i.e. they are in a state of relative starvation. Thus, the expression of genes in this state will presumably reflect those involved in bacterial survival strategies. The following results describe the differential gene expression observed between the different growth models, at different time points.

5.2.2.1 Biofilm (BF) versus planktonic exponential (PE) growth

BF grown cells were compared to planktonic cells extracted during exponential growth and genes expressed in greater amounts in the BF cells (Table 5.11) were identified. A group of genes belonging to the gamma haemolysin operon (*hlgABC*) were highly expressed (*hlgA* is written as *hlgII* in the table). These genes are known to function as two component cytolytic (pore-forming) toxins in the disruption and lysis of erythrocytes and leukocytes. In particular, the killing of leukocytes is a potentially important immune-evasion mechanism of the bacteria, which results in the weakening of the immune system of the host, and enables the bacteria to gain access to nutrients stored in leukocyte cells (Menestrina *et al* 2003). Two further toxins (beta haemolysin and exotoxin 6) were also highly expressed. Since these cells were grown over 7 days, the expression of such toxins would be expected. The fact that some of the toxins were related (*hlg* operon) provides confidence in the data observed. Furthermore, two further related genes, *splA* and *splB* (proteases) were also detected with greater expression in the BF cells. As noted previously, exo-enzymes are important for biofilm detachment following accumulation and growth.

Another group of genes, belonging to a single operon, expressed at higher levels in the BF phase were *bsaGEF* (MW1758-60). These encode a set of hypothetical proteins (carried on genomic island nuSaαβ) that are similar to the *EpiFEG* proteins of the gallidermin superfamily, specifically epidermin in this case. Uniprot and KEGG descriptions of the *bsa* genes describes their function as involved in ABC type transport systems. Epidermin is a lantibiotic (*lan*thionine containing an*tibiotic*), that is, a bacteriocin (ribosomally synthesized peptide antibiotic) containing lanthionine (or other modified amino acid) following post translational modification (Saris *et al.* 1996, Nissen-Mayer and Nes 1997). The *EpiFEG* transport system is involved in protecting lantibiotic producing bacteria against fatality associated with its own product. Although no known lantibiotic was detected in this data set, the presence of the *bsa* genes suggests this is a lantibiotic producing strain.

Type 1 *agr* transcripts including RNAIII were detected in both data sets but at higher levels in the BF cells. RNAIII is thought to be involved in the regulation of the haemolysin toxin genes as well as the proteases (Dunman *et al.*, 2001). Another pair of related genes highly expressed in the BF cells were *arcC* (carbamate kinase) and SA2424 (hypothetical protein similar to *arcR* transcriptional regulator according to Uniprot). The *arcR* regulator is known to positively regulate the expression of the *arcABDC* operon for arginine catabolism under aerobic conditions. Expression of *arcC* with a possible regulator indicates actively metabolic cells. Finally, of the classical surface-associated proteins involved in attachment, *clfA* was significantly more highly expressed in the BF cells.

Table 5.11 Biofilm-associated genes (BF v PE).

The log ratios indicate down regulation between BF and PE cells; that is, expression was higher in BF cells.

Gene ID	Product description	log_2 ratios (BFvPE)	TF
SA0857	hypothetical protein, similar to negative regulator of genetic competence MecA	1.94	4258.09
gi|78172212 (agr)	agr type 1 subset specific 1 12 16 17 20 25	1.96	8503.09
SAV2451	subfamily S9C non-peptidase homologues	2.49	4003.09
MW1798	glutamate ABC transporter ATP-binding protein	2.49	990.09
MW1959 :RNAIII	from agr (hld)	3.47	3541.09
MW1881:hlb	beta haemolysin	3.81	3648.59
SA2119	hypothetical protein, similar to dehydrogenase	3.81	1249.09
MW1759:bsaE	hypothetical protein, similar to EpiE (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	4.27	910.09
SA2424	hypothetical protein, similar to transcription regulator Crp/Fnr family protein	4.60	2501.59
MW1754:splB	serine protease	4.66	1070.59
MW1758:bsaG	hypothetical protein, similar to EpiG (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	4.97	1119.09
MW2343:hlgC	gamma-haemolysin component C	5.39	1476.09
SAV0422:set6	exotoxin 6	5.48	2308.09
MW2553:arc	carbamate kinase	5.49	9679.09
MW0764:clfA	fibrinogen-binding protein	5.53	6708.09
MW1760:bsaF	hypothetical protein, similar to EpiF (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	5.69	1469.09
MW2344:hlgB	gamma-haemolysin component B	6.16	4178.59
MW0396	hypothetical protein	6.24	4793.59
SAV1813:splA	subfamily S1B unassigned peptidases (exoprotein A)	6.40	1043.09
MW2342:hlgII	gamma haemolysin chain II	7.84	2900.09

Traditionally, biofilm-grown cells are known to have a heterogeneous phenotype. Anwar and colleagues (1992) found that the physiology of cells within a biofilm was determined by the position of individual cells within the multi-layered community. In particular, cells on the surface of the biofilm were larger in size and more metabolically active compared to the base layer cells, probably due to the easier access to oxygen and nutrients. Under these experimental conditions, cells grown in the CDFF instrument had constant access to oxygen and nutrients. This is dissimilar to a classic biofilm model in that cells were not allowed to accumulate. It seems these cells display a distinct phenotype different from rapidly growing planktonic cells and from relatively starved planktonic cells.

The list of genes with higher expression in the PE phase (Table 5.12) included many more hypothetical genes. In particular, several genes with regulatory (or putative regulatory) functions were detected (putative functions had been assigned to most by sequence homology). As before, the expression of co-ordinately regulated genes was detected. Firstly, sigma factor B (*sigB*) and SigB regulatory protein (*rsbU*) displayed similar expression levels. Sigma factor B is a known regulator of several virulence factors, which generally, exerts indirect regulation. Other related genes detected include bone sialoprotein binding proteins *sdrD* and *bbp* (a putative *sdrE* homolog), both adhesion proteins. Others genes in the same category (adhesins) were fibronectin binding protein (*fnbp*), *spaA* (IgG binding protein A) and *sbi* (Ig binding protein).

Table 5.12 PE associated genes (BF v PE)

Gene ID	Product description	log_2 ratios (BvPE)	TF
SA0641	conserved hypothetical protein, similar to transcriptional regulator	-1.96	11159.09
MW1280	ABC transporter (ATP-binding protein) homolog	-2.09	1009.09
SAV0526:radA	family S16 non-peptidase homologues	-2.18	1094.09
MW1988:sigB	sigma factor B	-2.24	1515.09
MW0681	hypothetical protein, similar to anion-binding protein	-2.25	2367.09
bbp	sdrE homolog (bone sialoprotein binding protein) O24	-2.26	1755.09
MW2341:sbi	IgG binding protein	-2.48	2010.09
MW0736:tpi	triosephosphate isomerase	-2.49	2722.09
SAV1044	family S33 unassigned peptidases	-2.52	1501.59
MW1991:rsbU	sigmaB regulation protein RsbU	-2.57	951.09
MW0339	hypothetical protein, similar to GTP-binding protein	-2.82	4257.59
SA0704	conserved hypothetical protein	-3.05	1726.09
MW0079	hypothetical protein similar to transmembrane efflux pump protein	-3.10	1152.59
MW2013	lipoprotein precursor	-3.13	6182.09
SA2421	hypothetical protein, similar to transcriptional regulator	-3.29	955.09
MW0283	hypothetical protein, similar to branched-chain amino acid uptake carrier	-3.44	1186.09
MW2421:fnb	fibronectin-binding protein homolog (for fnbA)	-3.46	1679.09
MW2217	secretory antigen precursor SsaA homolog	-3.55	1255.59
SA1949	lytic regulatory protein truncated with Tn554 (truncated-SA)	-3.85	1678.09
MW0342:ssb	single-strand DNA-binding protein of phage phi Sa 2mw	-3.89	23821.09
MW0517:sdrD	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	-4.25	1071.09
MW0595	hypothetical protein, similar to ABC transporter ATP-binding protein	-4.31	2083.09
MW0084:spa	IgG binding protein A	-4.68	13121.09
SA0726:gapR	glycolytic operon regulator	-5.09	8642.09
MW2336	hypothetical protein, similar to ABC transporter, periplasmic amino acid-binding protein	-6.19	7108.09
SAV0433:set15	exotoxin 15	-6.87	5847.09

5.2.2.2 BF versus Planktonic Stationary (PS) growth

The RNA preparations from cells cultured overnight were dominated by structural RNA and contained relatively little mRNA. This result is unsurprising since PS cells are less metabolically active than other cells examined in this study. When BF and PS transcripts were compared (Table 5.13) the list of genes expressed at higher levels in the BF cells was similar to that described in the BF versus PE comparison (Table 5.11). In particular the *hlg* operon genes, *hlb*, exotoxin 6 and SA0857 were still highly expressed in BF cells. Other genes not listed in Table 5.11 were also detected due to lower levels of transcripts being present in the PS cells. Firstly, two important regulators *saeR* and *lexA*, showed high expression levels. *LexA* is an SOS regulatory protein, whilst *saeR* is the regulator of the Sae locus which is known to be essential for *hlb* transcription (Giraudo *et al.*, 1997) detected here. Furthermore, *ftsH* (SAV0511), a cell division protein, was also detected. Others included elastin binding protein (MW1369) and a hypothetical protein associated with fibrinogen binding protein function (MW1041). The latter are surface-associated proteins (MSCRAMMs) that are expected to be expressed during active growth. The BF expression of *ftsH* is further evidence of the presence of actively growing cells. These data suggest the cells grown within the CDFF model have an actively growing phenotype.

The gene list for PE cells (Table 5.14) only contained two genes, a hypothetical protein with no associated function and a gene found in type IV SCC*mec*. The detection of only two genes expressed at levels higher than those found in the BF cells is an indication of the low mRNA content of the PS cells.

Table 5.13 Biofilm-associated genes (BF v PS)

Gene ID	Product description	log_2 ratios (BvPS)	TF
MW0986	conserved hypothetical protein	2.09	3199.70
MW1760:bsaF	hypothetical protein, similar to EpiF (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	2.15	1054.70
SA0641	conserved hypothetical protein, similar to transcriptional regulator	2.21	1162.70
SA1174:LexA	SOS regulatory LexA protein	2.57	2197.70
MW1758:bsaG	hypothetical protein, similar to EpiG (Genomic island nuSaβ2) [putative lantibiotic ABC transporter protein]	2.58	1017.20
MW1369:epbS	elastin binding protein	3.11	1645.70
SA0661:saeR	response regulator	3.23	1164.20
SAV0511:ftsH	FtsH-2 peptidase	3.32	1837.70
SA0857	hypothetical protein, similar to negative regulator of genetic competence MecA	3.78	3310.70
MW1041	hypothetical protein, similar to fibrinogen-binding protein	4.01	1044.70
MW1834	hypothetical protein, similar to ferritin	4.39	1700.70
MW2343:hlgC	gamma-haemolysin component C	4.43	1348.70
MW2342:hlgII	gamma haemolysin chain II	4.54	2164.20
SAV0422:set6	exotoxin 6	4.80	1912.70
MW1881:hlb	beta haemolysin	4.87	2392.70
MW2344:hlgB	gamma-haemolysin component B	5.07	2593.70
MW0396	hypothetical protein	5.83	2929.20

Table 5.14 PS associated genes (BF v PS)

Gene ID	Product description	log_2 ratios (BvPS)	TF
TypeIVb-F	SCC*mec* IVb	-2.62	1653.70
MW1744	hypothetical protein	-2.88	1162.20

5.2.2.3 PE versus PS growth

PE and PS cells were compared in order to explore growth phase dependent differences in this strain. The results are summarised in Table 5.15 (higher expression in PS cells) and Table 5.16 (higher expression in PE cells). A few interesting trends were observed from which some conclusions may be drawn. Of the PS associated genes described in Table 5.15, 3 of the genes (marked with *) were also expressed at higher levels in the biofilm-associated cells described in Table 5.11. Furthermore, of the genes expressed at higher levels in the PE cells (Table 5.16), 4 were also seen in the BF cells (Table 5.13). These results indicate that BF cells grown on the CDFF instrument display a typical phenotype with differences from both the actively growing cells (PE) and the stationary (PS) cells. The CDFF therefore appears to be a helpful model system for investigating natural biofilms. Furthermore, this model is amenable to exploitation in studies designed to improve our understanding of the mechanisms underlining biofilm survival and regulation after long periods of growth.

Table 5.15 PS associated genes (PE v PS)

Gene ID	Product description	log_2 ratios (PEvPS)	TF
SAV2451	subfamily S9C non-peptidase homologues	-1.96	1822.85
MW1744	hypothetical protein	-2.21	1236.85
MW0764:clfA *	fibrinogen-binding protein	-3.88	1188.85
MW2553:arcC *	carbamate kinase	-5.29	3889.85
E16-0463 - SAR0424	Bactrltoxin	-5.76	1682.35
SA2424 *	hypothetical protein, similar to transcription regulator Crp/Fnr family protein	-11.24	1335.10

Table 5.16 PE genes (PE v PS)

Gene ID	Product description	log_2 ratios (PEvPS)	TF
MW0543:pta	phosphotransacetylase	1.99	1200.85
MW0085:sarH1	staphylococcal accessory regulator A homologue	2.07	910.85
MW0681	hypothetical protein, similar to anion-binding protein	2.12	1720.85
MW1942	hypothetical protein similar to leukocidin chain lukM precursor	2.24	1001.85
MW0518:sdrE	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	2.32	2128.35
MW1834	hypothetical protein, similar to ferritin	2.38	981.35
SAV1044	family S33 unassigned peptidases	2.42	917.35
MW1092:gmk	guanylate kinase	2.70	1682.85
MW2341:sbi	IgG binding protein	3.03	1158.85
SAV0520	family C56 non-peptidase homologues	3.04	980.85
MW0339	hypothetical protein, similar to GTP-binding protein	3.04	3652.85
SAV0743:pepT	peptidase T	3.11	985.35
MW0595	hypothetical protein, similar to ABC transporter ATP-binding protein	3.48	940.35
MW0736:tpi	triosephosphate isomerase	3.62	1564.85
SA0641	conserved hypothetical protein, similar to transcriptional regulator	3.63	7489.85
MW2013	lipoprotein precursor	3.67	3001.85
SA0661:saeR	response regulator	4.10	1327.85
SA0726:gapR	glycolytic operon regulator	4.27	6134.85
MW2336	hypothetical protein, similar to ABC transporter, periplasmic amino acid-binding protein	4.45	5965.85
MW0342:ssb	single-strand DNA-binding protein of phage phi Sa 2mw	4.52	19214.85
MW0084:spa	IgG binding protein A	4.59	7604.85
SAV0511:ftsH	FtsH-2 peptidase	4.64	4535.85
SA1949	lytic regulatory protein truncated with Tn554 (truncated-SA)	5.06	1026.85

5.2.3 Biofilm summary

From the point of view of exploring differences in the virulence-associated genes of biofilm versus planktonic cells, the following observations were noted. Both Table 5.11 and Table 5.13 list genes expressed at higher levels in BF. Nine genes appear in both tables (the raw data are summarised in Table 5.17) and six (all toxins) were highly expressed in both comparisons including BF cells. The remaining three genes were not consistently highly expressed between the BF RNA preparations. Based on these data high level expression of the haemolysin genes was the strongest feature of the biofilm state.

Table 5.17 Biofilm-specific gene list

Gene ID	Product Description	log_2 BvPE	TF	log_2 BvPS	TF	log_2 PEvPS	TF
MW0396	hypothetical protein	6.24	4793.59	5.83	2929.20	0.00	1.00
MW2344:hlgB	gamma-haemolysin component B	6.16	4178.59	5.07	2593.70	0.00	1.00
MW1881:hlb	beta haemolysin	3.81	3648.59	4.87	2392.70	-1.29	0.41
MW2342:hlgII	gamma haemolysin chain II	7.84	2900.09	4.54	2164.20	0.00	1.00
SAV0422:set6	exotoxin 6	5.48	2308.09	4.80	1912.70	0.00	1.00
MW2343:hlgC	gamma-haemolysin component C	5.39	1476.09	4.43	1348.70	0.00	1.00
MW1760:bsaF	hypothetical protein, similar to EpiF (Genomic island nuSaβ2)	5.69	1469.09	2.15	1054.70	-3.18	0.11
MW1758:bsaG	hypothetical protein, similar to EpiG (Genomic island nuSaβ2)	4.97	1119.09	2.58	1017.20	-0.79	0.58
SA0857	hypothetical protein, similar to negative regulator of genetic competence MecA	1.94	4258.09	3.78	3310.70	1.94	3.83

CHAPTER 6.0 CONCLUSION

Staphylococcus aureus is the most clinically significant member of the *Staphylococcus* genus. Its ability to be associated with asymptomatic carriage as well as cause a wide range of infections, from superficial to more life-threatening, has prompted much research into its genotypic and phenotypic characteristics. Furthermore, interest in its ability to maintain a colonist nature in addition to 'switching' to a virulent phenotype in epidemiologically diverse niches has partly driven the sequencing of >20 genomes of various genetic backgrounds representative of the most successful lineages. These initiatives have provided data for comparative studies (i.e. sequence comparisons, PCR analysis and microarray investigation) that have helped researchers to understand further the ever-changing evolutionary landscape of this organism and in particular the potential virulence- and fitness-factors that may contribute to its success. In this context, microarray analysis provides a rapid method by which bacterial genomes can be interrogated.

This thesis details the development and use of a partial composite *S. aureus* microarray for the interrogation of *S. aureus* isolates for putative virulence-associated genes. Protocols were subsequently developed for comparative genome hybridisation (CGH) and transcription profiling studies. The microarray was used to investigate a panel of internationally recognised pandemic and sporadic lineages of healthcare-associated (HA-) and community-associated (CA-) *S. aureus* strains to provide insights into variation within and between lineages. The microarray provides a comprehensive tool by which these lineages can be either clustered or differentiated based on genetic markers. This work supports a clonal population structure for this organism, as described previously (Enright, 2000). The data presented show congruence with MLST as a method for defining lineages, with the advantage of providing information on the virulence-gene content of each test isolate. Yet, while MLST focuses on a set of core conserved genes, this work

demonstrated similar outcomes based on either core or accessory genes, indicating that accessory genome components are partially conserved within and between lineages. Profiles specific to each of the main clonal groups could be easily identified and so the microarray could form the basis of a more informative approach for epidemiological typing of clinical isolates in future.

The array was used to explore genotypic differences that may account for variations in epidemiology and pathogenicity of recognised successful, epidemic, or pandemic strains. In particular, epidemic MRSA (EMRSA; defined as "*those which have been identified in two or more patients in two or more hospitals*"; O'Neill *et al.*, 2001) were interrogated for features that may provide insights into traits important in the success, transmission and/or pathogenicity of these lineages. The NCTC collection of UK epidemic MRSA strains, EMRSA-1 to EMRSA-17 (excluding EMRSA-6), were analysed using the array. Features unique to a particular group of strains could be detected and selected from the array. However, analysis of the array data did not implicate any specific virulence-associated gene with the epidemicity potential of these strains. Based on the array data, the success of any particular clone appears to result from the complex interplay between factors. With respect to the important UK epidemic MRSA strains, EMRSA-15/-16, success of these clones might be attributed to the acquisition of extra accessory genes, and in particular the ΦSa2mw genes (section 4.1.2.3). Alternatively, it is possible that non-virulence factors (not represented on the array) may contribute to the epidemicity of these strains.

The microarray data were also used to probe for recombination events that may have occurred during the evolution of successful *S. aureus* lineages. Genetic exchange is central to the evolution of any organism. The hypothesis was that large-scale recombination events may have had an impact on the way epidemic strains survive in hospital environments and colonise the human nasopharynx. Such events may only be detected if they are relatively

recent in evolutionary terms since they are obscured by subsequent mutations and exchanges. A particular large-scale recombination appears to have been involved in the origin of the EMRSA-15 lineage (ST22-SCC*mec*IV), the most prevalent epidemic HA-MRSA in the UK. This work supports the hypothesis that CC22, which includes the EMRSA-15 lineage, evolved via a large scale recombination between CC8-like and CC30-like strains.

Notably, the array data also show that the ST772 strain (a member of CC1 by MLST) is an interesting example of clonal diversification. Distance trees built by simple matching of the individual probe reactions revealed that the ST772 strain clustered with an ST59 strain (USA1000) and not with other members of CC1. This result, which was supported by other laboratory data, emphasises the importance of using multiple genetic loci for genotypic characterisation and highlights the limitation of MLST in focussing purely on a limited number of conserved genetic loci. ST772 is proposed to be an intermediary strain in the evolution of the CC59 lineage from a common ST1 ancestor (section 4.2.4).

A further objective of this work was to explore patterns of genotypic variation that might account for differences in epidemiology and pathogenicity of the HA and CA *S. aureus* strains. A set of marker genes (16 consecutive genes of the bacteriophage ΦSa3 locus) associated with phage structural units linked with the CA phenotype were identified. These genes were identified in 10 of the 16 CA strains, but only in 1 of the HA strains. A recent study by Goerke and colleagues (2009) showed that *S. aureus* bacteriophage morphology genes correlate with phage serogroup type (A, B, Fa or Fb). By extrapolation of these data, the morphology related genes of the ΦSa3 locus identified in this study represent phage serogroup type Fb. Similarly, the ΦN315 morphology genes on the array represent the Fa serotype, whilst the ΦSa2mw genes (also present on the array) represent the A type. The only morphotype not identified by the array was the B serotype.

However, *in silico* analysis of the sequenced strains showed that the B serotype was exclusive to the HA-strains. The absence of B serotype phage in CA strains was also confirmed with the recently sequenced CA strains from this study (ST772, WA-MRSA, ST866, USA1000 and the Queensland clone ST93).

Goerke and colleagues (2009) also showed that the A serogroup phage comprised of two divisions, those that carried the *pvl* genes (PVL phage) and those that did not. Additionally, PVL phage were also of the Fb serotype. These PVL associated A/Fb phage were seen only in the CA strains, whilst the non PVL-associated A group phage were seen in only two of the sequenced HA strains (and the two CA MSSA strains). However, the non PVL-associated Fb phage were not present in any of the HA strains. That is to say, since the genes recognised by the array predominantly indicate the presence of an A/Fb serogroup phage in the CA strains (or possibly a mosaic phage within these divisions, which would be possible via exchange of modules according to the theory of modular evolution), it is believed that the carriage of PVL-associated phage (but not *pvl* itself, since it is absent from some of the CA strains), is necessary for the CA phenotype. It is possible that a mosaic phage that combines the fitness factors of the *pvl*-carrying A/ Fb phage and the *sak/ chp/ scn* carrying Fb phage could produce a virulent phenotype in those strains carrying phage of this morphotype. Presently, these genes represent a partial marker for the CA phenotype. This finding further supports the role of the accessory genome, and in particular bacteriophage, in the epidemiology of *S. aureus*.

The virulence-associated gene microarray was also used to investigate gene expression in transcription profiling studies. As an analytic technique, transcriptomic analysis provides more detailed information about bacterial response to its microenvironment. In the first transcription study, the virulence gene expression profiles of cells of an *agr* mutant strain (SH1001) and its unmodified parental strain (SH1000) were compared to determine the

effect of this mutation on virulence gene regulation. SH1000 was derived from the commonly used genetic lineage *S. aureus* 8325-4 (RN6390) and a functional *rsbU* gene (8325-4 *rsbU*$^{+}$) inserted. RsbU, the positive regulator of Sigma factor B, is often non-functional in strain 8325-4 due to a mutation (8325-4 *rsbU*).

The *agr* locus of *S. aureus* encodes a regulatory RNA molecule (RNAIII) known to influence the transcription of many virulence-associated genes. The consensus model is that genes encoding surface-associated proteins are generally expressed in the early exponential phase during cell proliferation (e.g. fibronectin binding protein A), whilst secreted proteins (e.g. α-haemolysin) are repressed. Conversely, in the stationary phase the opposite regulation effect has been reported. Strains were analysed for changes in expression levels between the early exponential- (5hr), late exponential- (7.5hr) and stationary- (overnight) phases. Cells were grown under the same conditions, on the same day, and extractions prepared in the same batch. Although the results demonstrate the anticipated involvement of the Agr operon in gene regulation, the mechanism involving RNAIII-related quorum sensing (usually triggered during cellular growth) was not evident. The observed changes in expression were apparently unrelated to the intracellular concentration of RNAIII since this transcript (although present at relatively high levels) was not induced or repressed. Nevertheless, RNAIII involvement could not be ruled out. It may be that the non-response of RNAIII to cell density may be an adapted feature of this common laboratory strain. For example, constant expression of RNAIII could be a feature that favours *in vitro* growth, conferring an advantage status to this strain in this environment such that factors involved in cellular proliferation and attachment (controlled via RNAIII) are continuously expressed instead of following traditional trends of expression. If this was the case, it would seem that the re-introduction of *rsbU* to strain NCTC 8325-4 (SH1000) was insufficient to re-establish normal environmental responses. Sabersheikh and Saunders (2004) in studies of the virulence-associated gene transcripts of

S. aureus also noted lack of a clear pattern of RNAIII transcription. Fluctuations in RNAIII levels were observed, and thought to be correlated with irregular growth patterns (also observed in this study). Sabersheikh and Saunders proposed that the mechanism by which RNAIII is induced is more complex than originally anticipated and that variations in this mechanism exist between different *S. aureus* clonotypes. It is therefore possible that continuous expression of RNAIII is a feature of this genetic background, and consequently rendering it an unsuitable model for investigating *agr*-mediated virulence gene regulation.

Finally, cells grown under biofilm simulating conditions (in the CDFF instrument) were used in comparative transcription studies against their planktonic counterparts to determine putative biofilm-associated virulence genes and also to determine the suitability of this model for further studies. Biofilm formation is a feature of *S. aureus* infections associated with implanted medical devices, especially within the hospital inpatient community. These challenging microbial communities are resistant to host defence mechanisms and respond inefficiently to standard antibiotics (Ceri *et al.* 1999, Donlan *et al.* 2002, Leid *et al.* 2002). Biofilm development is thought to be controlled via gene expression that differs significantly between biofilm-associated and planktonic organisms (Beloin and Ghigo, 2005). In this study, *S. aureus* cells grown in a continuous culture biofilm were analysed in parallel with the same strain grown in suspension. Biofilm growth was supported over a seven day period in the CDFF instrument, whilst the planktonic cells were grown continuously for 21 hours, with sampling at 5 hours (early exponential phase) and 21 hours (stationary phase).

The description and analysis of the differential gene expression observed between the different growth models, revealed the predominant feature of the biofilm state to be high level expression of the haemolysin genes. These genes are known to function as pore-forming toxins involved in the disruption and lysis of erythrocytes and leukocytes, an

important immune-evasion mechanism which results in the weakening of the immune system of the host, and enables the bacteria to gain access to nutrients stored in leukocyte cells (Menestrina *et al.*, 2003). The CDFF instrument appeared to be a helpful model system for investigating natural biofilms. Biofilm cells displayed a typical phenotype different from both the actively growing planktonic exponential cells and the planktonic stationary cells. This model therefore is amenable to exploitation in studies designed to improve our understanding of the mechanisms underlining biofilm survival and regulation after long periods of growth.

CHAPTER 7.0 FUTURE WORK

Microarrays offer great potential; the information obtained from such experiments far exceeds that of any multiplex PCR. As a data mining tool, strain/ species-specific features can be more easily determined, reducing hands-on time. Informative array design and interpretation is highly dependent on genome sequencing and annotation data. *S. aureus* represents one of the most sequenced organisms, with >20 sequenced genomes (including several unpublished genomes described in this thesis).

- The primary objective in any future work will involve updating the array features to include genes from more recently sequenced genomes.
- Additionally, designing customised arrays in a multiarray format would provide a rapid method for screening for differential markers.

This project revealed the potential of phage serogroup typing to distinguish HA and CA *S. aureus* strains. Within the current array design, only three (A, Fa and Fb) of the four known serogroups could be identified. The remaining B serogroup could only be sought via *in silico* analysis. The results showed that PVL-associated phage of the A and Fb serogroups are strongly associated with CA status, whilst the B serogroup was seen only in HA strains.

- Future work would entail designing further oligos for all known phage serotypes. Bacteriophage that integrate at the Sa2 and Sa3 loci in particular have been found in most of the sequenced strains to date, indicating their potential involvement in *S. aureus* success.

- The investigation of this same phenomenon with respect to pathogenicity islands (PI) or genomic islands (GI) is warranted. The majority of virulence-associated factors and antimicrobial resistance determinants are components of the accessory genome, and have been mapped to bacteriophage, PI and GI. The ability to detect and characterise these mobile genetic elements in strains that have not been sequenced (and in particular in clinical isolates), will greatly enhance our understanding of the virulence potential of each strain.

The approach of characterising the carriage of groups of mobile genetic elements may provide insights into recombination events between the various groups. By designing probes unique to each Fb serogroup phage (and also other serogroups), recombination events may be described with greater certainty.

The results described in this thesis indicate a novel phage taxon could be associated with CA status. If so, identifying phage types associated with CA status should lead to the more precise identification of genetic elements associated with CA infections. Thus, by pinpointing these genetic elements, control measures and even therapeutics to more effectively combat these infections may be developed.

The array was successfully used to study RNA transcripts to monitor changes in gene expression. Analysis of the commonly used SH1000/SH1001 strains in *agr* studies revealed them to be an interesting model. RNAIII-related quorum sensing, a key feature of virulence-regulation, was not evident. Insertion of a functional *rsbU* gene (8325-4 *rsbU*$^{+}$) into the commonly used genetic lineage *S. aureus* 8325-4 (RN6390) did not lead to reconstruction of the expected functionality of the Agr operon.

- Future work would entail obtaining additional model strains for the study of *agr*-mediated virulence regulation.

Finally, the array was used to study cells grown under biofilm simulating conditions in the CDFF instrument. The biofilm grown cells displayed a characteristic phenotype different from planktonic cells. This model may be exploited further in studies designed to improve our understanding of the mechanisms underlining biofilm survival and regulation after long periods of growth.

- Future studies may seek to monitor the expression profiles of biofilm cells grown on simulated tissues. For example, growing the cells on a layer simulating the human basal lamina should increase our understanding of biofilm attachment processes.

CHAPTER 8.0 REFERENCES

Aires de Sousa, M. & De Lencastre, H. 2003, "Evolution of sporadic isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) in hospitals and their similarities to isolates of community-acquired MRSA", *J Clin Microbiol*, vol. 41, no. 8, pp. 3806-15.

Aires de Sousa, M. & De Lencastre, H. 2004, "Bridges from hospitals to the laboratory: Genetic portraits of methicillin-resistant *Staphylococcus aureus* clones", *Fems Immunology and Medical Microbiology*, vol. 40, no. 2, pp. 101-111.

Aligholi, M., Emaneini, M., Jabalameli, F., Shahsavan, S., Dabiri, H., & Sedaght, H. 2008, "Emergence of high-level vancomycin-resistant *Staphylococcus aureus* in the Imam Khomeini hospital in Tehran", *Medical Principles and Practice*, vol. 17, no. 5, pp. 432-434.

Anonymous 1959. "Staphylococcal disease in hospitals", *Lancet*, vol.1, pp.185-186.

Anonymous 1999. From the Centres for Disease Control and Prevention. Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus* – Minnesota and North Dakota, 1997-1999. *JAMA,* vol. 282, pp 1123-1125.

Antignac, A. & Tomasz, A. 2009, "Reconstruction of the phenotypes of methicillin-resistant *Staphylococcus aureus* by replacement of the staphylococcal cassette chromosome *mec* with a plasmid-borne copy of *Staphylococcus sciuri* pbpD gene", *Antimicrobial Agents and Chemotherapy*, vol. 53, no. 2, pp. 435-441.

Anwar, H., Strap, J. L., & Costerton, J. W. 1992, "Susceptibility of biofilm cells of *Pseudomonas aeruginosa* to bactericidal actions of whole blood and serum", *FEMS Microbiol Lett*, vol. 71, no. 3, pp. 235-241.

Appelbaum, P. C. 2007, "Microbiology of antibiotic resistance in *Staphylococcus aureus*", *Clinical Infectious Diseases*, vol. 45, pp. S165-S170.

Arbeit R.D. 1997, Laboratory procedures for epidemiological analysis, pp253-286. *In* K. B. Crossley and G. L. Archer (ed.), The staphylococci in human disease. Churchill Livingstone, New York, NY.

Archer, G. L., Karchmer, A. W., Vishniavsky, N., & Johnston, J. L. 1984, "Plasmid-pattern analysis for the differentiation of infecting from noninfecting *Staphylococcus-Epidermidis*", *Journal of Infectious Diseases*, vol. 149, no. 6, pp. 913-920.

Arvidson, S. 2000. Extracellular enzymes, pp. 379-385. *In* V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (ed.), Gram-positive pathogens. ASM Press, Washington, D.C.

Arvidson, S. & Tegmark, K. 2001, "Regulation of virulence determinants in *Staphylococcus aureus*", *International Journal of Medical Microbiology*, vol. 291, no. 2, pp. 159-170.

Aucken, H. M., Ganner, M., Murchan, S., Cookson, B. D., & Johnson, A. P. 2002, "A new UK strain of epidemic methicillin-resistant *Staphylococcus aureus* (EMRSA-17)

resistant to multiple antibiotics", *Journal of Antimicrobial Chemotherapy*, vol. 50, no. 2, pp. 171-175.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., & Hiramatsu, K. 2002, "Genome and virulence determinants of high virulence community-acquired MRSA", *Lancet*, vol. 359, no. 9320, pp. 1819-1827.

Baba, T., Kuwahara-Arai, K., Uchiyama, I., Takeuchi, F., Ito, T., & Hiramatsu, K. 2009, "Complete genome sequence of *Macrococcus caseolyticus* strain JSCS5402, reflecting the ancestral genome of the human-pathogenic Staphylococci", *Journal of Bacteriology*, vol. 191, no. 4, pp. 1180-1190.

Baird-Parker, A. C. 1990, "The staphylococci: an introduction", *Soc Appl Bacteriol Symp Ser*, vol. 19, pp. 1S-8S.

Balaban, N. & Novick, R. P. 1995, "Autocrine regulation of toxin synthesis by *Staphylococcus aureus*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 5, pp. 1619-1623.

Balaban, N., Goldkorn, T., Gov, Y., Hirshberg, M., Koyfman, N., Matthews, H. R., Nhan, R. T., Singh, B., & Uziel, O. 2001, "Regulation of *Staphylococcus aureus* pathogenesis via target of RNAIII-activating protein (TRAP)", *Journal of Biological Chemistry*, vol. 276, no. 4, pp. 2658-2667.

Bayer, M. G., Heinrichs, J. H., & Cheung, A. L. 1996, "The molecular architecture of the sar locus in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 178, no. 15, pp. 4563-4570.

Bazinet, C. & King, J. 1985, "The DNA translocating vertex of double-stranded bacteriophage", *Ornston, L.N.(Ed.) Annual Review of Microbiology, Vol.39.Xii+727P.Annual Reviews Inc.: Palo Alto, Calif., Usa.Illus* pp. 109-130.

Becker, K., Harmsen, D., Mellmann, A., Meier, C., Schumann, P., Peters, G., & von Eiff, C. 2004, "Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of Staphylococcus species", *Journal of Clinical Microbiology*, vol. 42, no. 11, pp. 4988-4995.

Becker, P., Hufnagle, W., Peters, G., & Herrmann, M. 2001, "Detection of differential gene expression in biofilm-forming versus planktonic populations of *Staphylococcus aureus* using micro-representational-difference analysis", *Applied and Environmental Microbiology*, vol. 67, no. 7, pp. 2958-2965.

Beenken, K. E., Blevins, J. S., & Smeltzer, M. S. 2003, "Mutation of *sarA* in *Staphylococcus aureus* limits biofilm formation", *Infection and Immunity*, vol. 71, no. 7, pp. 4206-4211.

Beenken, K. E., Dunman, P. M., McAleese, F., Macapagal, D., Murphy, E., Projan, S. J., Blevins, J. S., & Smeltzer, M. S. 2004, "Global gene expression in *Staphylococcus aureus* biofilms", *Journal of Bacteriology*, vol. 186, no. 14, pp. 4665-4684.

Beloin, C. & Ghigo, J. M. 2005, "Finding gene-expression patterns in bacterial biofilms", *Trends in Microbiology*, vol. 13, no. 1, pp. 16-19.

Benito, Y., Lina, G., Greenland, T., Etienne, J., & Vandenesch, F. 1998, "Trans-complementation of a *Staphylococcus aureus agr* mutant by *Staphylococcus lugdunensis agr* RNAIII", *Journal of Bacteriology*, vol. 180, no. 21, pp. 5780-5783.

Benito, Y., Kolb, F. A., Romby, P., Lina, G., Etienne, J., & Vandenesch, F. 2000, "Probing the structure of RNAIII, the *Staphylococcus aureus agr* regulatory RNA, and identification of the RNA domain involved in repression of protein A expression", *RNA-A Publication of the RNA Society*, vol. 6, no. 5, pp. 668-679.

Bens, C. C. P. M., Voss, A., & Klaassen, C. H. W. 2006, "Presence of a novel DNA methylation enzyme in methicillin-resistant *Staphylococcus aureus* isolates associated with pig farming leads to uninterpretable results in standard pulsed-field gel electrophoresis analysis", *Journal of Clinical Microbiology*, vol. 44, no. 5, pp. 1875-1876.

Berger-Bachi, B. & Rohrer, S. 2002, "Factors influencing methicillin resistance in Staphylococci", *Archives of Microbiology*, vol. 178, no. 3, pp. 165-171.

Bischoff, M., Entenza, J. M., & Giachino, P. 2001, "Influence of a functional *sigB* operon on the global regulators *sar* and *agr* in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 183, no. 17, pp. 5171-5179.

Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D., Murphy, E., Mounts, W., Berger-Bachi, B., & Projan, S. 2004, "Microarray-based analysis of the *Staphylococcus aureus* sigma(B) regulon", *Journal of Bacteriology*, vol. 186, no. 13, pp. 4085-4099.

Black, L. W. 1989, "DNA packaging in DNA bacteriophages", *Annual Review of Microbiology*, vol. 43, pp. 267-292.

Blevins, J. S., Gillaspy, A. F., Rechtin, T. M., Hurlburt, B. K., & Smeltzer, M. S. 1999, "The staphylococcal accessory regulator (*sar*) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (*cna*) in an *agr*-independent manner", *Molecular Microbiology* , vol. 33, no. 2, pp. 317-326.

Blevins, J. S., Beenken, K. E., Elasri, M. O., Hurlburt, B. K., & Smeltzer, M. S. 2002, "Strain-dependent differences in the regulatory roles of *sarA* and *agr* in *Staphylococcus aureus*", *Infection and Immunity*, vol. 70, no. 2, pp. 470-480.

Boles, B. R., Thoendel, M., & Singh, P. K. 2004, "Self-generated diversity produces "insurance effects" in biofilm communities", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 47, pp. 16630-16635.

Botstein, D. 1980, "A theory of modular evolution for bacteriophages", *Ann N Y Acad Sci*, vol. 354, pp. 484-490.

Boving, M. K., Pedersen, L. N., & Moller, J. K. 2009, "Eight-plex PCR and liquid-array detection of bacterial and viral pathogens in cerebrospinal fluid from patients with suspected meningitis", *Journal of Clinical Microbiology*, vol. 47, no. 4, pp. 908-913.

Boyd, E. F., Davis, B. M., & Hochhut, B. 2001, "Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria", *Trends in Microbiology*, vol. 9, no. 3, pp. 137-144.

Bronner, S., Stoessel, P., Gravet, A., Monteil, H., & Prevost, G. 2000, "Variable expressions of *Staphylococcus aureus* bicomponent leucotoxins semiquantified by competitive reverse transcription-PCR", *Applied and Environmental Microbiology*, vol. 66, no. 9, pp. 3931-3938.

Bronner, S., Monteil, H., & Prevost, G. 2004, "Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications", *Fems Microbiology Reviews*, vol. 28, no. 2, pp. 183-200.

Broudy, T. B., Pancholi, V., & Fischetti, V. A. 2001, "Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells", *Infection and Immunity*, vol. 69, no. 3, pp. 1440-1443.

Brunskill, E. W. & Bayles, K. W. 1996, "Identification and molecular characterization of a putative regulatory locus that affects autolysis in *Staphylococcus aureus*", *Journal of Bacteriology* , vol. 178, no. 3, pp. 611-618.

Brussow, H. & Desiere, F. 2001, "Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages", *Molecular Microbiology*, vol. 39, no. 2, pp. 213-222.

Brussow, H., Canchaya, C., & Hardt, W. D. 2004, "Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion", *Microbiology and Molecular Biology Reviews*, vol. 68, no. 3, pp. 560-602.

Bunney, W. E., Bunney, B. G., Vawter, M. P., Tomita, H., Li, J., Evans, S. J., Choudary, P. V., Myers, R. M., Jones, E. G., Watson, S. J., & Akil, H. 2003, "Microarray technology: A review of new strategies to discover candidate vulnerability genes in psychiatric disorders", *American Journal of Psychiatry*, vol. 160, no. 4, pp. 657-666.

Burroughs, N. J., Marsh, P., & Wellington, E. M. H. 2000, "Mathematical analysis of growth and interaction dynamics of streptomycetes and a bacteriophage in soil", *Applied and Environmental Microbiology*, vol. 66, no. 9, pp. 3868-3877.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., & Brussow, H. 2003, "Prophage genomics", *Microbiology and Molecular Biology Reviews*, vol. 67, no. 2, pp. 238-276.

Carpentier, B. & Cerf, O. 1993, "Biofilms and their consequences, with particular reference to hygiene in the food-industry", *Journal of Applied Bacteriology*, vol. 75, no. 6, pp. 499-511.

Carter, B., Wu, G. H., Woodward, M. J., & Anjum, M. F. 2008, "A process for analysis of microarray comparative genomics hybridisation studies for bacterial genomes", *BMC Genomics*, vol. 9, no. 53.

Casjens, S., Wyckoff, E., Hayden, M., Sampson, L., Eppler, K., Randall, S., Moreno, E. T., & Serwer, P. 1992, "Bacteriophage-P22 portal protein is part of the gauge that regulates packing density of intravirion DNA", *Journal of Molecular Biology*, vol. 224, no. 4, pp. 1055-1074.

Cassat, J., Dunman, P. M., Murphy, E., Projan, S. J., Beenken, K. E., Palm, K. J., Yang, S. J., Rice, K. C., Bayles, K. W., & Smeltzer, M. S. 2006, "Transcriptional profiling of a *Staphylococcus aureus* clinical isolate and its isogenic *agr* and *sarA* mutants reveals global differences in comparison to the laboratory strain RN6390", *Microbiology*, vol. 152, pp. 3075-3090.

Causton, H. C., Quackenbush, J., Brazma, A., Causton, H. C., Quackenbush, J., & Brazma, A. 2003, "Microarray gene expression data analysis: A beginner's guide", *Microarray gene expression data analysis: A beginner's guide* p. vi-160.

Ceri, H., Olson, M. E., Stremick, C., Read, R. R., Morck, D., & Buret, A. 1999, "The Calgary Biofilm Device: New technology for rapid determination of antibiotic susceptibilities of bacterial biofilms", *Journal of Clinical Microbiology*, vol. 37, no. 6, pp. 1771-1776.

Chamberlain, N. R. & Imanoel, B. 1996, "Genetic regulation of fatty acid modifying enzyme from *Staphylococcus aureus*", *Journal of Medical Microbiology*, vol. 44, no. 2, pp. 125-129.

Chambers, H. F., Hartman, B. J., & Tomasz, A. 1985, "Increased amounts of a novel penicillin-binding protein in a strain of methicillin-resistant *Staphylococcus aureus* exposed to Nafcillin", *Journal of Clinical Investigation*, vol. 76, no. 1, pp. 325-331.

Chambers, H. F. 2001, "The changing epidemiology of *Staphylococcus aureus*?", *Emerging Infectious Diseases*, vol. 7, no. 2, pp. 178-182.

Chambers, H. F. & Deleo, F. R. 2009, "Waves of resistance: *Staphylococcus aureus* in the antibiotic era", *Nature Reviews Microbiology*, vol. 7, no. 9, pp. 629-641.

Chan, P. F. & Foster, S. J. 1998, "The role of environmental factors in the regulation of virulence-determinant expression in *Staphylococcus aureus* 8325-4", *Microbiology-Uk*, vol. 144, pp. 2469-2479.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., & Fodor, S. P. A. 1996, "Accessing genetic information with high-density DNA arrays", *Science*, vol. 274, no. 5287, pp. 610-614.

Cheung, A. L., Koomey, J. M., Butler, C. A., Projan, S. J., & Fischetti, V. A. 1992, "Regulation of exoprotein expression in *Staphylococcus aureus* by a locus (Sar) Distinct from *agr*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 14, pp. 6462-6466.

Cheung, A. L. & Projan, S. J. 1994, "Cloning and sequencing of *sarA* of *Staphylococcus aureus*, a gene required for the expression of *agr*", *Journal of Bacteriology*, vol. 176, no. 13, pp. 4168-4172.

Cheung, A. L., Bayer, M. G., & Heinrichs, J. H. 1997, "Sar genetic determinants necessary for transcription of RNAII and RNAIII in the *agr* locus of *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 179, no. 12, pp. 3963-3971.

Cheung, A. L., Bayer, A. S., Zhang, G. Y., Gresham, H., & Xiong, Y. Q. 2004, "Regulation of virulence determinants in vitro and in vivo in *Staphylococcus aureus*", *Fems Immunology and Medical Microbiology*, vol. 40, no. 1, pp. 1-9.

Chien, Y. T. & Cheung, A. L. 1998, "Molecular interactions between two global regulators, *sar* and *agr*, in *Staphylococcus aureus*", *Journal of Biological Chemistry*, vol. 273, no. 5, pp. 2645-2652.

Chien, Y. T., Manna, A. C., Projan, S. J., & Cheung, A. L. 1999, "*SarA*, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for sar-dependent gene regulation", *Journal of Biological Chemistry*, vol. 274, no. 52, pp. 37169-37176.

Chongtrakool, P., Ito, T., Ma, X. X., Kondo, Y., Trakulsomboon, S., Tiensasitorn, C., Chavalit, T., Song, J. H., & Hiramatsu, K. 2006, "Staphylococcal cassette chromosome *mec* (SCC*mec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCC*mec* elements", *Antimicrobial Agents and Chemotherapy*, vol. 50, no. 3, pp. 1001-1012.

Christianson, S., Golding, G. R., Campbell, J., Mulvey, M. R., & Canadian, N., I. 2007, "Comparative genomics of Canadian epidemic lineages of methicillin-resistant *Staphylococcus aureus*", *Journal of Clinical Microbiology*, vol. 45, no. 6, pp. 1904-1911.

Cockfield, J. D., Pathak, S., Edgeworth, J. D., & Lindsay, J. A. 2007, "Rapid determination of hospital-acquired meticillin-resistant *Staphylococcus aureus* lineages", *Journal of Medical Microbiology*, vol. 56, no. 5, pp. 614-619.

Coia, J. E., Noorhussain, I., & Platt, D. J. 1988, "Plasmid profiles and restriction enzyme fragmentation patterns of plasmids of methicillin-sensitive and methicillin-resistant isolates of *Staphylococcus aureus* from hospital and the community", *Journal of Medical Microbiology*, vol. 27, no. 4, pp. 271-276.

Cookson, B. D., Robinson, D. A., Monk, A. B., Murchan, S., Deplano, A., de Ryck, R., Struelens, M. J., Scheel, C., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., Cuny, C., Witte, W., Tassios, P. T., Legakis, N. J., van Leeuwen, W., van Belkum, A., Vindel, A., Garaizar, J., Haeggman, S., Olsson-Liljequist, B., Ransjo, U., Muller-Premru, M., Hryniewicz, W., Rossney, A., O'Connell, B., Short, B. D., Thomas, J., O'Hanlon, S., & Enright, M. C. 2007, "Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: The HARMONY collection", *Journal of Clinical Microbiology*, vol. 45, no. 6, pp. 1830-1837.

Corkill, J. E., Anson, J. J., Griffiths, P., & Hart, C. A. 2004, "Detection of elements of the staphylococcal cassette chromosome (SCC) in a methicillin-susceptible (*mecA* gene negative) homologue of a fucidin-resistant MRSA", *Journal of Antimicrobial Chemotherapy*, vol. 54, no. 1, pp. 229-231.

Costerton, J. W., Stewart, P. S., & Greenberg, E. P. 1999, "Bacterial biofilms: A common cause of persistent infections", *Science*, vol. 284, no. 5418, pp. 1318-1322.

Couto, I., deLencastre, H., Severina, E., Kloos, W., Webster, J. A., Hubner, R. J., Sanches, I. S., & Tomasz, A. 1996, "Ubiquitous presence of a *mecA* homologue in natural isolates of *Staphylococcus sciuri*", *Microbial Drug Resistance-Mechanisms Epidemiology and Disease* , vol. 2, no. 4, pp. 377-391.

Couto, I., Sanches, I. S., Sa-Leao, R., & De Lencastre, H. 2000, "Molecular characterization of *Staphylococcus sciuri* strains isolated from humans", *Journal of Clinical Microbiology*, vol. 38, no. 3, pp. 1136-1143.

Cox, R. A., Conquest, C., Mallaghan, C., & Marples, R. R. 1995, "A major outbreak of methicillin-resistant *Staphylococcus aureus* caused by a new phage-type (EMRSA-16)", *Journal of Hospital Infection*, vol. 29, no. 2, pp. 87-106.

Cramton, S. E., Gerke, C., Schnell, N. F., Nichols, W. W., & Gotz, F. 1999, "The intercellular adhesion (ica) locus is present in *Staphylococcus aureus* and is required for biofilm formation", *Infection and Immunity*, vol. 67, no. 10, pp. 5427-5433.

Crisostomo, M. I., Westh, H., Tomasz, A., Chung, M., Oliveira, D. C., & De Lencastre, H. 2001, "The evolution of methicillin resistance in *Staphylococcus aureus*: Similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 17, pp. 9865-9870.

Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I., & Penades, J. R. 2001, "Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation", *Journal of Bacteriology*, vol. 183, no. 9, pp. 2888-2896.

Dassy, B., Hogan, T., Foster, T. J., & Fournier, J. M. 1993, "Involvement of the accessory gene regulator (*agr*) in expression of type-5 capsular polysaccharide by *Staphylococcus aureus*", *Journal of General Microbiology*, vol. 139, pp. 1301-1306.

Daugherty, S. & Low, M. G. 1993, "Cloning, expression, and mutagenesis of phosphatidylinositol-specific phospholipase-C from *Staphylococcus aureus* - a potential Staphylococcal virulence factor", *Infection and Immunity*, vol. 61, no. 12, pp. 5078-5089.

David, M. D., Kearns, A. M., Gossain, S., Ganner, M., & Holmes, A. 2006, "Community-associated meticillin-resistant *Staphylococcus aureus*: nosocomial transmission in a neonatal unit", *Journal of Hospital Infection*, vol. 64, no. 3, pp. 244-250.

De Lencastre, H., Wu, S. W., Pinho, M. G., Ludovice, A. M., Filipe, S., Gardete, S., Sobral, R., Gill, S., Chung, M., & Tomasz, A. 1999, "Antibiotic resistance as a stress response: Complete sequencing of a large number of chromosomal loci in *Staphylococcus aureus* strain COL that impact on the expression of resistance to methicillin", *Microbial Drug Resistance-Mechanisms Epidemiology and Disease*, vol. 5, no. 3, pp. 163-175.

De Lencastre, H., Chung, M., & Westh, H. 2000, "Archaic strains of methicillin-resistant *Staphylococcus aureus*: Molecular and microbiological properties of isolates from the 1960s in Denmark", *Microbial Drug Resistance-Mechanisms Epidemiology and Disease*, vol. 6, no. 1, pp. 1-10.

De Lencastre, H., Oliveira, D., & Tomasz, A. 2007, "Antibiotic resistant *Staphylococcus aureus*: a paradigm of adaptive power", *Current Opinion in Microbiology*, vol. 10, no. 5, pp. 428-435.

De Lencastre, H., Dejonge, B. L. M., Matthews, P. R., & Tomasz, A. 1994, "Molecular aspects of methicillin resistance in *Staphylococcus aureus*", *Journal of Antimicrobial Chemotherapy*, vol. 33, no. 1, pp. 7-24.

Deresinski, S. 2005, "Methicillin-resistant *Staphylococcus aureus*: An evolutionary, epidemiologic, and therapeutic odyssey", *Clinical Infectious Diseases*, vol. 40, no. 4, pp. 562-573.

Derisi, J. L., Iyer, V. R., & Brown, P. O. 1997, "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, vol. 278, no. 5338, pp. 680-686.

Desiere, F., Lucchini, S., & Brussow, H. 1999, "Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate cos-site Streptococcus thermophilus bacteriophage Sfi21", *Virology*, vol. 260, no. 2, pp. 244-253.

Desiere, F., Pridmore, R. D., & Brossow, H. 2000, "Comparative genomics of the late gene cluster from Lactobacillus phages", *Virology*, vol. 275, no. 2, pp. 294-305.

Deurenberg, R. H. & Stobberingh, E. E. 2008, "The evolution of *Staphylococcus aureus*", *Infection Genetics and Evolution*, vol. 8, no. 6, pp. 747-763.

Deurenberg, R. H. & Stobberingh, E. E. 2009, "The molecular evolution of hospital- and community-associated methicillin-resistant *Staphylococcus aureus*", *Current Molecular Medicine*, vol. 9, no. 2, pp. 100-115.

Diep, B. A., Gill, S. R., Chang, R. F., Phan, T. H., Chen, J. H., Davidson, M. G., Lin, F., Lin, J., Carleton, H. A., Mongodin, E. F., Sensabaugh, G. F., & Perdreau-Remington, F. 2006, "Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*", *Lancet*, vol. 367, no. 9512, pp. 731-739.

Diep, B. A. & Otto, M. 2008, "The role of virulence determinants in community-associated MRSA pathogenesis", *Trends in Microbiology*, vol. 16, no. 8, pp. 361-369.

Donahue, J. A. & Baldwin, J. N. 1966, "Haemolysin and leukocidin production by 80/81 strains of *Staphylococcus aureus*", *Journal of Infectious Diseases*, vol. 116, no. 3, p. 324.

Donlan, R. M. & Costerton, J. W. 2002, "Biofilms: Survival mechanisms of clinically relevant microorganisms", *Clinical Microbiology Reviews*, vol. 15, no. 2, pp. 167-193.

Doskar, J., Pallova, P., Pantucek, R., Rosypal, S., Ruzickova, V., Pantuckova, P., Kailerova, J., Kleparnik, K., Mala, Z., & Bocek, P. 2000, "Genomic relatedness of *Staphylococcus aureus* phages of the International Typing Set and detection of serogroup A, B, and F prophages in lysogenic strains", *Canadian Journal of Microbiology*, vol. 46, no. 11, pp. 1066-1076.

Dove, W. F. 1971 "Biological inferences", pp 297-3 12. *In* A. D. HERSHEY (ed). The Bacteriophage Lambda, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Dufour, P., Gillet, Y., Bes, M., Lina, G., Vandenesch, F., Floret, D., Etienne, J., & Richet, H. 2002, "Community-acquired methicillin-resistant *Staphylococcus aureus* infections in France: Emergence of a single clone that produces Panton-Valentine leukocidin", *Clinical Infectious Diseases*, vol. 35, no. 7, pp. 819-824.

Dunman, P. M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., Brown, E. L., Zagursky, R. J., Shlaes, D., & Projan, S. J. 2001, "Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci", *Journal of Bacteriology*, vol. 183, no. 24, pp. 7341-7353.

Dunman, P. M., Mounts, W., McAleese, F., Immermann, F., Macapagal, D., Marsilio, E., McDougal, L., Tenover, F. C., Bradford, P. A., Petersen, P. J., Projan, S. J., & Murphy, E. 2004, "Uses of *Staphylococcus aureus* GeneChips in genotyping and genetic composition analysis", *Journal of Clinical Microbiology*, vol. 42, no. 9, pp. 4275-4283.

Dunne, W. M. 2002, "Bacterial adhesion: Seen any good biofilms lately?", *Clinical Microbiology Reviews*, vol. 15, no. 2, pp. 155-166.

Ellington MJ, Ganner M, Warner M, Cookson BD, Kearns AM. 2009, "Polyclonal multiply antibiotic-resistant methicillin-resistant *Staphylococcus aureus* with Panton-Valentine Leucocidin in England". *J Antimicrob Chemother*, doi: 10.1093/jac/dkp386.

Ellison, D. W. & Miller, V. L. 2006, "Regulation of virulence by members of the MarR/SlyA family", *Current Opinion in Microbiology*, vol. 9, no. 2, pp. 153-159.

Ender, M., Berger-Bachi, B., & McCallum, N. 2009, "A novel DNA-binding protein modulating methicillin resistance in *Staphylococcus aureus*", *BMC Microbiology*, vol. 9. no. 15.

Enright, M.C. 2008, "The population structure of *Staphylococcus aureus*", pp.29-44. *In* J Lindsay (ed.), Staphylococcus molecular genetics. Caister Academic Press, Norfolk, UK.

Enright, M. C., Day, N. P. J., Davies, C. E., Peacock, S. J., & Spratt, B. G. 2000, "Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*", *Journal of Clinical Microbiology*, vol. 38, no. 3, pp. 1008-1015.

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., & Spratt, B. G. 2002, "The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA)", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 11, pp. 7687-7692.

Eppler, K., Wyckoff, E., Goates, J., Parr, R., & Casjens, S. 1991, "Nucleotide-sequence of the bacteriophage-P22 genes required for DNA packaging", *Virology*, vol. 183, no. 2, pp. 519-538.

Feil, E. J., Smith, J. M., Enright, M. C., & Spratt, B. G. 2000, "Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data", *Genetics*, vol. 154, no. 4, pp. 1439-1450.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., & Spratt, B. G. 2004, "eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data", *Journal of Bacteriology*, vol. 186, no. 5, pp. 1518-1530.

Feng, Y., Chen, C. J., Su, L. H., Hu, S., Yu, J., & Chiu, C. H. 2008, "Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics", *Fems Microbiology Reviews*, vol. 32, no. 1, pp. 23-37.

Ferry, T., Perpoint, T., Vandenesch, F., & Etienne, J. 2005, "Virulence determinants in *Staphylococcus aureus* and their involvement in clinical syndromes", *Curr Infect Dis Rep*, vol. 7, no. 6, pp. 420-428.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R., & Musser, J. M. 2001, "Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 15, pp. 8821-8826.

Fokkens, W. J. & Scheeren, R. A. 2000, "Upper airway defence mechanisms", *Paediatr Respir Rev*, vol. 1, no. 4, pp. 336-341.

Foster, T. J. & Hook, M. 1998, "Surface protein adhesins of *Staphylococcus aureus*", *Trends in Microbiology*, vol. 6, no. 12, pp. 484-488.

Fournier, B. 2008, "Global regulators of *Staphylococcus aureus*", pp.131-184. *In* J Lindsay (ed.), Staphylococcus molecular genetics. Caister Academic Press, Norfolk, UK.

Fournier, B. & Hooper, D. C. 2000, "A new two-component regulatory system involved in adhesion autolysis, and extracellular proteolytic activity of *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 182, no. 14, pp. 3955-3964.

Fournier, B., Klier, A., & Rapoport, G. 2001, "The two-component system ArlS-ArlR is a regulator of virulence gene expression in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 41, no. 1, pp. 247-261.

Fournier, B. & Klier, A. 2004, "Protein A gene expression is regulated by DNA supercoiling which is modified by the ArlS-ArlR two-component system of *Staphylococcus aureus*", *Microbiology*, vol. 150, pp. 3807-3819.

Fowler, V. G., Fey, P. D., Reller, L. B., Chamis, A. L., Corey, G. R., & Rupp, M. E. 2001, "The intercellular adhesin locus ica is present in clinical isolates of *Staphylococcus aureus* from bacteremic patients with infected and uninfected prosthetic joints", *Medical Microbiology and Immunology*, vol. 189, no. 3, pp. 127-131.

Francois, P., Harbarth, S., Huyghe, A., Renzi, G., Bento, M., Gervaix, A., Pittet, D., & Schrenzel, J. 2008, "Methicillin-resistant *Staphylococcus aureus*, Geneva, Switzerland, 1993-2005", *Emerging Infectious Diseases*, vol. 14, no. 2, pp. 304-307.

Francois, P. and Schrenzel, J. 2008, Rapid diagnosis and typing of *Staphylococcus aureus*, p.71-88. *In* J Lindsay (ed.), Staphylococcus molecular genetics. Caister Academic Press, Norfolk, UK.

Frees, D., Qazi, S. N. A., Hill, P. J., & Ingmer, H. 2003, "Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence", *Molecular Microbiology*, vol. 48, no. 6, pp. 1565-1578.

Frees, D., Chastanet, A., Qazi, S., Sorensen, K., Hill, P., Msadek, T., & Ingmer, H. 2004, "Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 54, no. 5, pp. 1445-1462.

Frees, D., Sorensen, K., & Ingmer, H. 2005, "Global virulence regulation in *Staphylococcus aureus*: Pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network", *Infection and Immunity*, vol. 73, no. 12, pp. 8100-8108.

Garvis, S., Mei, J. M., Ruiz-Albert, J., & Holden, D. W. 2002, "*Staphylococcus aureus* svrA: a gene required for virulence and expression of the agr locus", *Microbiology*, vol. 148, pp. 3235-3243.

Gaskill, M. E. & Khan, S. A. 1988, "Regulation of the Enterotoxin-B Gene in *Staphylococcus aureus*", *Journal of Biological Chemistry*, vol. 263, no. 13, pp. 6276-6280.

Geisinger, E., Adhikari, R. P., Jin, R. Z., Ross, H. F., & Novick, R. P. 2006, "Inhibition of rot translation by RNAIII, a key feature of agr function", *Molecular Microbiology*, vol. 61, no. 4, pp. 1038-1048.

Gerke, C., Kraft, A., Sussmuth, R., Schweitzer, O., & Gotz, F. 1998, "Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the Staphylococcus epidermidis polysaccharide intercellular adhesin", *Journal of Biological Chemistry*, vol. 273, no. 29, pp. 18586-18593.

Gillespi, D. & Spiegelm, S. 1965, "A Quantitative Assay for DNA-RNA Hybrids with Dna Immobilized on A Membrane", *Journal of Molecular Biology*, vol. 12, no. 3, pp. 829-829.

Gilot, P., Lina, G., Cochard, T., & Poutrel, B. 2002, "Analysis of the genetic variability of genes encoding the RNA III-activating components Agr and TRAP in a population of *Staphylococcus aureus* strains isolated from cows with mastitis", *Journal of Clinical Microbiology*, vol. 40, no. 11, pp. 4060-4067.

Gilot, P. & van Leeuwen, W. 2004, "Comparative analysis of agr locus diversification and overall genetic variability among bovine and human *Staphylococcus aureus* isolates", *Journal of Clinical Microbiology*, vol. 42, no. 3, pp. 1265-1269.

Giraudo, A. T., Cheung, A. L., & Nagel, R. 1997, "The sae locus of *Staphylococcus aureus* controls exoprotein synthesis at the transcriptional level", *Archives of Microbiology*, vol. 168, no. 1, pp. 53-58.

Goerke, C., Kummel, M., Dietz, K., & Wolz, C. 2003, "Evaluation of intraspecies interference due to *agr* polymorphism in *Staphylococcus aureus* during infection and colonization", *Journal of Infectious Diseases*, vol. 188, no. 2, pp. 250-256.

Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., Broker, B. M., Doskar, J., & Wolz, C. 2009, "Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages", *Journal of Bacteriology*, vol. 191, no. 11, pp. 3462-3468.

Gomes, A. R., Westh, H., & De Lencastre, H. 2006, "Origins and evolution of methicillin-resistant *Staphylococcus aureus* clonal lineages", *Antimicrobial Agents and Chemotherapy*, vol. 50, no. 10, pp. 3237-3244.

Gordon, R. J. & Lowy, F. D. 2008, "Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection", *Clinical Infectious Diseases*, vol. 46, pp. S350-S359.

Groom, A. V., Wolsey, D. H., Naimi, T. S., Smith, K., Johnson, S., Boxrud, D., Moore, K. A., & Cheek, J. E. 2001, "Community-acquired methicillin-resistant *Staphylococcus aureus* in a rural American Indian community", *Journal of the American Medical Association*, vol. 286, no. 10, pp. 1201-1205.

Hallin, M., Denis, O., Deplano, A., de Ryck, R., Crevecoeur, S., Rottiers, S., de Mendonca, R., & Struelens, M. J. 2008, "Evolutionary relationships between sporadic and epidemic strains of healthcare-associated methicillin-resistant *Staphylococcus aureus*", *Clinical Microbiology and Infection*, vol. 14, no. 7, pp. 659-669.

Han, S. H., Chin, B. S., Lee, H. S., Jeong, S. J., Choi, H. K., Kim, C. K., Kim, C. O., Yong, D., Choi, J. Y., Song, Y. G., Lee, K., & Kim, J. M. 2009, "Recovery of both vancomycin-resistant Enterococci and methicillin-resistant *Staphylococcus aureus* from culture of a single clinical specimen from colonized or infected patients", *Infection Control and Hospital Epidemiology*, vol. 30, no. 2, pp. 130-138.

Hanssen, A. M., Kjeldsen, G., & Sollid, J. U. E. 2004, "Local variants of staphylococcal cassette chromosome *mec* in sporadic methicillin-resistant *Staphylococcus aureus* and methicillin-resistant coagulase-negative Staphylococci: Evidence of horizontal gene transfer?", *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 1, pp. 285-296.

Hanssen, A. M. & Sollid, J. U. E. 2007, "Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus aureus* and coaaulase-neizative stahylococci from Norway", *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 5, pp. 1671-1677.

Hardiman, G. 2004, "Microarray platforms - comparisons and contrasts", *Pharmacogenomics*, vol. 5, no. 5, pp. 487-502.

Harmsen, D., Claus, H., Witte, W., Rothganger, J., Claus, H., Turnwald, D., & Vogel, U. 2003, "Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management", *Journal of Clinical Microbiology*, vol. 41, no. 12, pp. 5442-5448.

Hartman, B. J. & Tomasz, A. 1984, "Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 158, no. 2, pp. 513-516.

Heilmann, C., Schweitzer, O., Gerke, C., Vanittanakom, N., Mack, D., & Gotz, F. 1996, "Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*", *Molecular Microbiology*, vol. 20, no. 5, pp. 1083-1091.

Heinrichs, J. H., Bayer, M. G., & Cheung, A. L. 1996, "Characterization of the *sar* locus and its interaction with agr in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 178, no. 2, pp. 418-423.

Hendrix, R. W. 2003, "Bacteriophage genomics", *Current Opinion in Microbiology*, vol. 6, no. 5, pp. 506-511.

Ilеusser, R., Ender, M., Berger-Bachi, B., & McCallum, N. 2007, "Mosaic staphylococcal cassette chromosome *mec* containing two recombinase loci and a new *mec* complex, B2", *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 1, pp. 390-393.

Higuchi, W., Takano, T., Teng, L. J., & Yamamoto, T. 2008, "Structure and specific detection of staphylococcal cassette chromosome mec type VII", *Biochemical and Biophysical Research Communications*, vol. 377, no. 3, pp. 752-756.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T., & Tenover, F. C. 1997, "Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility", *Journal of Antimicrobial Chemotherapy*, vol. 40, no. 1, pp. 135-136.

Hiramatsu, K., Cui, L., Kuroda, M., & Ito, T. 2001, "The emergence and evolution of methicillin-resistant *Staphylococcus aureus*", *Trends in Microbiology*, vol. 9, no. 10, pp. 486-493.

Hiramatsu, K., Katayama, Y., Yuzawa, H., & Ito, T. 2002, "Molecular genetics of methicillin-resistant *Staphylococcus aureus*", *International Journal of Medical Microbiology*, vol. 292, no. 2, pp. 67-74.

Hiramatsu, K., Watanabe, S., Takeuchi, F., Ito, T., & Baba, T. 2004, "Genetic characterization of methicillin-resistant *Staphylococcus aureus*", *Vaccine*, vol. 22, p. S5-S8.

Holden, M. T. G., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P. J., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G., & Parkhill, J. 2004, "Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9786-9791.

Holden M. and Lindsay J. 2008, Whole genomes: Sequence, microarray and systems biology, pp.1-28. *In* J Lindsay (ed.), Staphylococcus molecular genetics. Caister Academic Press, Norfolk, UK.

Holmberg, S. D., Wachsmuth, I. K., Hickmanbrenner, F. W., & Cohen, M. L. 1984, "Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing S*almonella typhimurium* isolates from outbreaks", *Journal of Clinical Microbiology*, vol. 19, no. 2, pp. 100-104.

Holt J. G. 1994, *Group 17. "Gram-positive cocci"*, pp. 527–558. *In* J.G. Holt, N.R. Krieg, P.H. Sneath, J.T. Staley and S.T. Williams (ed.), *Bergey's Manual of Determinative Bacteriology*, Williams and Wilkins, Baltimore.

Horsburgh, M. J., Aish, J. L., White, I. J., Shaw, L., Lithgow, J. K., & Foster, S. J. 2002, "sigma(B) modulates virulence determinant expression and stress resistance: Characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325-4", *Journal of Bacteriology*, vol. 184, no. 19, pp. 5457-5467.

Iandolo, J. J., Worrell, V., Groicher, K. H., Qian, Y. D., Tian, R. Y., Kenton, S., Dorman, A., Ji, H. G., Lin, S. P., Loh, P., Qi, S. L., Zhu, H., & Roe, B. A. 2002, "Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325", *Gene*, vol. 289, no. 1-2, pp. 109-118.

Ingavale, S. S., van Wamel, W., & Cheung, A. L. 2003, "Characterization of RAT, an autolysis regulator in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 48, no. 6, pp. 1451-1466.

Inglis, B., Waldron, H., & Stewart, P. R. 1987, "Molecular relatedness of *Staphylococcus aureus* typing phages measured by DNA hybridization and by high-resolution thermal-denaturation analysis", *Archives of Virology*, vol. 93, no. 1-2, pp. 69-80.

Ito, T., Katayama, Y., & Hiramatsu, K. 1999, "Cloning and nucleotide sequence determination of the entire mec DNA of pre-methicillin-resistant *Staphylococcus aureus* N315", *Antimicrobial Agents and Chemotherapy*, vol. 43, no. 6, pp. 1449-1458.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., & Hiramatsu, K. 2001, "Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 45, no. 5, pp. 1323-1336.

Ito, T., Okuma, K., Ma, X. X., Yuzawa, H., & Hiramatsu, K. 2003, "Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: genomic island SCC", *Drug Resistance Updates*, vol. 6, no. 1, pp. 41-52.

Ito, T., Ma, X. X., Takeuchi, F., Okuma, K., Yuzawa, H., & Hiramatsu, K. 2004, "Novel type v staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, ccrC", *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 7, pp. 2637-2651.

Izard, N. C., Hachler, H., Grehn, M., & Kayser, F. H. 1992, "Ribotyping of coagulase-negative Staphylococci with special emphasis on intraspecific typing of *Staphylococcus epidermidis*", *Journal of Clinical Microbiology*, vol. 30, no. 4, pp. 817-823.

Jansen, W. T. M., Beitsma, A. A., Koeman, C. J., van Wamel, W. J. B., Verhoef, J., & Fluit, A. C. 2006, "Novel mobile variants of staphylococcal cassette chromosome *mec* in *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 50, no. 6, pp. 2072-2078.

Janzon, L., Lofdahl, S., & Arvidson, S. 1986, "Evidence for a coordinate transcriptional control of alpha-toxin and protein-a synthesis in *Staphylococcus aureus*", *Fems Microbiology Letters*, vol. 33, no. 2-3, pp. 193-198.

Jarraud, S., Lyon, G. J., Figueiredo, A. M. S., Gerard, L., Vandenesch, F., Etienne, J., Muir, T. W., & Novick, R. P. 2000, "Exfoliatin-producing strains define a fourth *agr* specificity group in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 182, no. 22, pp. 6517-6522.

Jarraud, S., Mougel, C., Thioulouse, J., Lina, G., Meugnier, H., Forey, F., Nesme, X., Etienne, J., & Vandenesch, F. 2002, "Relationships between *Staphylococcus aureus* genetic background, virulence factors, *agr* groups (Alleles), and human disease", *Infection and Immunity*, vol. 70, no. 2, pp. 631-641.

Jensen, S. O. & Lyon, B. R. 2009, "Genetics of antimicrobial resistance in *Staphylococcus aureus*", *Future Microbiology*, vol. 4, no. 5, pp. 565-582.

Jevons, M. P., Rolinson, G. N., & Knox, R. 1961, "Celbenin-resistant Staphylococci", *British Medical Journal*, vol. 1, no. 521, pp. 124-125

Ji, G. Y., Beavis, R., & Novick, R. P. 1997, "Bacterial interference caused by autoinducing peptide variants", *Science*, vol. 276, no. 5321, pp. 2027-2030.

Johnson, A. P., Aucken, H. M., Cavendish, S., Ganner, M., Wale, M. C. J., Warner, M., Livermore, D. M., & Cookson, B. D. 2001, "Dominance of EMRSA-15 and-16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS)", *Journal of Antimicrobial Chemotherapy*, vol. 48, no. 1, pp. 143-144.

Jones, M. N. 2005, "Use of liposomes to deliver bactericides to bacterial biofilms", *Methods Enzymol.*, vol. 391, pp. 211-228.

Kahl, B. C., Becker, K., Friedrich, A. W., Clasen, J., Sinha, B., von Eiff, C., & Peters, G. 2003, "*agr*-Dependent bacterial interference has no impact on long-term colonization of *Staphylococcus aureus* during persistent airway infection of cystic fibrosis patients", *Journal of Clinical Microbiology*, vol. 41, no. 11, pp. 5199-5201.

Kaito, C., Omae, Y., Matsumoto, Y., Nagata, M., Yamaguchi, H., Aoto, T., Ito, T., Hiramatsu, K., & Sekimizu, K. 2008, "A Novel Gene, *fudoh*, in the SCC*mec* Region Suppresses the Colony Spreading Ability and Virulence of *Staphylococcus aureus*", *Plos One*, vol. 3, no. 12, doi: 10.1371/journal.pone.0003921.

Kaneko, J., Kimura, T., Narita, S., Tomita, T., & Kamio, Y. 1998, "Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phi PVL-carrying Panton-Valentine leukocidin genes", *Gene*, vol. 215, no. 1, pp. 57-67.

Karlsson, A., Saravia-Otten, P., Tegmark, K., Morfeldt, E., & Arvidson, S. 2001, "Decreased amounts of cell wall-associated protein A and fibronectin-binding proteins in *Staphylococcus aureus sarA* mutants due to up-regulation of extracellular proteases", *Infection and Immunity*, vol. 69, no. 8, pp. 4742-4748.

Karlsson, A. & Arvidson, S. 2002, "Variation in extracellular protease production among clinical isolates of *Staphylococcus aureus* due to different levels of expression of the protease repressor *sarA*", *Infection and Immunity*, vol. 70, no. 8, pp. 4239-4246.

Katayama, Y., Ito, T., & Hiramatsu, K. 2000, "A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 6, pp. 1549-1555.

Katayama, Y., Ito, T., & Hiramatsu, K. 2001, "Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: Role of IS431-mediated mecI deletion in expression of resistance in mecA-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*", *Antimicrobial Agents and Chemotherapy*, vol. 45, no. 7, pp. 1955-1963.

Katayama, Y., Zhang, H. Z., Hong, D., & Chambers, H. F. 2003, "Jumping the barrier to beta-lactam resistance in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 185, no. 18, pp. 5465-5472.

Katayama, Y., Robinson, D. A., Enright, M. C., & Chambers, H. F. 2005, "Genetic background affects stability of *mecA* in *Staphylococcus aureus*", *Journal of Clinical Microbiology*, vol. 43, no. 5, pp. 2380-2383.

Kearns, A. M., Turner, A. J. L., Eltringham, G. J. A., & Freeman, R. 2002, "Rapid detection and quantification of CMV DNA in urine using LightCycler (TM)-based real-time PCR", *Journal of Clinical Virology*, vol. 24, no. 1-2, pp. 131-134.

Kennedy, A. D., Otto, M., Braughton, K. R., Whitney, A. R., Chen, L., Mathema, B., Mediavilla, J. R., Byrne, K. A., Parkins, L. D., Tenover, F. C., Kreiswirth, B. N., Musser, J. M., & Deleo, F. R. 2008, "Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: Recent clonal expansion and diversification", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, pp. 1327-1332.

Khodakov, D. A., Zakharova, N. V., Gryadunov, D. A., Filatov, F. P., Zasedatelev, A. S., & Mikhailovich, V. M. 2008, "An oligonucleotide microarray for multiplex real-time PCR identification of HIV-1HBV and HCV", *Biotechniques*, vol. 44, no. 2, pp. 241-246.

Klevens, R. M., Morrison, M. A., Fridkin, S. K., Reingold, A., Petit, S., Gershman, K., Ray, S., Harrison, L. H., Lynfield, R., Dumyati, G., Townes, J. M., Craig, A. S., Fosheim, G., McDougal, L. K., & Tenover, F. C. 2006, "Community-associated methicillin-resistant *Staphylococcus aureus* and healthcare risk factors", *Emerging Infectious Diseases*, vol. 12, no. 12, pp. 1991-1993.

Kloos, W. E. 1980, "Natural-populations of the genus Staphylococcus", *Annual Review of Microbiology*, vol. 34, pp. 559-592.

Kloos, W. E. 1990, "Systematics and the natural history of staphylococci 1", *Soc Appl Bacteriol Symp Ser*, vol. 19, pp. 25S-37S.

Kondo, Y., Ito, T., Ma, X. X., Watanabe, S., Kreiswirth, B. N., Etienne, J., & Hiramatsu, K. 2007, "Combination of multiplex PCRs for staphylococcal cassette chromosome *mec* type assignment: Rapid identification system for *mec*, *ccr*, and major differences in junkyard regions", *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 1, pp. 264-274.

Kreiswirth, B., Kornblum, J., Arbeit, R. D., Eisner, W., Maslow, J. N., Mcgeer, A., Low, D. E., & Novick, R. P. 1993, "Evidence for a clonal origin of methicillin resistance in *Staphylococcus-aureus*", *Science*, vol. 259, no. 5092, pp. 227-230.

Kullik, I., Giachino, P., & Fuchs, T. 1998, "Deletion of the alternative sigma factor sigma(B) in *Staphylococcus aureus* reveals its function as a global regulator of virulence genes", *Journal of Bacteriology*, vol. 180, no. 18, pp. 4814-4820.

Kumamoto, C. A. & Vinces, M. D. 2005, "Contributions of hyphae and hypha-co-regulated genes to Candida albicans virulence", *Cellular Microbiology*, vol. 7, no. 11, pp. 1546-1554.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L. Z., Oguchi, A., Aoki, K., Nagai, Y., Lian, J. Q., Ito, T., Kanamori, M., Matsumaru, H.,

Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., & Hiramatsu, K. 2001, "Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*", *Lancet*, vol. 357, no. 9264, pp. 1225-1240.

Kuroda, M., Kuroda, H., Oshima, T., Takeuchi, F., Mori, H., & Hiramatsu, K. 2003, "Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 49, no. 3, pp. 807-821.

Kuroda, M., Yamashita, A., Hirakawa, H., Kumano, M., Morikawa, K., Higashide, M., Maruyama, A., Inose, Y., Matoba, K., Toh, H., Kuhara, S., Hattori, M., & Ohta, T. 2005, "Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 37, pp. 13272-13277.

Kwon, N. H., Park, K. T., Moon, J. S., Jung, W. K., Kim, S. H., Kim, J. M., Hong, S. K., Koo, H. C., Joo, Y. S., & Park, Y. H. 2005, "Staphylococcal cassette chromosome *mec* (SCC*mec*) characterization and molecular analysis for methicillin-resistant *Staphylococcus aureus* and novel SCC*mec* subtype IVg isolated from bovine milk in Korea", *Journal of Antimicrobial Chemotherapy*, vol. 56, no. 4, pp. 624-632.

Leclercq, R. 2002, "Staphylococci resistant to antibiotic therapy", *Annales Francaises D Anesthesie et de Reanimation*, vol. 21, no. 5, pp. 375-383.

Lee, J. S. & Stewart, P. R. 1985, "The virion proteins and ultrastructure of *Staphylococcus aureus* bacteriophages", *Journal of General Virology*, vol. 66, no. pt9, pp. 2017-2027.

Leid, J. G., Shirtliff, M. E., Costerton, J. W., & Stoodley, P. 2002, "Human leukocytes adhere to, penetrate, and respond to *Staphylococcus aureus* biofilms", *Infection and Immunity*, vol. 70, no. 11, pp. 6339-6345.

Le arrec, C., vanSinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., & Fayard, B. 1997, "Two groups of bacteriophages infecting Streptococcus thermophilus can be distinguished on the basis of model of packaging and genetic determinants for major structural proteins", *Applied and Environmental Microbiology*, vol. 63, no. 8, pp. 3246-3253.

Li, M., Diep, B. A., Villaruz, A. E., Braughton, K. R., Jiang, X. F., Deleo, F. R., Chambers, H. F., Lu, Y., & Otto, M. 2009, "Evolution of virulence in epidemic community-associated methicillin-resistant *Staphylococcus aureus*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 14, pp. 5883-5888.

Li, R. G., Manna, A. C., Dai, S. D., Cheung, A. L., & Zhang, G. Y. 2003, "Crystal structure of the SarS protein from *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 185, no. 14, pp. 4219-4225.

Lina, B., Bes, M., Vandenesch, F., Greenland, T., Etienne, J., & Fleurette, J. 1993, "Role of bacteriophages in genomic variability of related coagulase-negative Staphylococci", *Fems Microbiology Letters*, vol. 109, no. 2-3, pp. 273-277.

Lina, G., Jarraud, S., Ji, G. Y., Greenland, T., Pedraza, A., Etienne, J., Novick, R. P., & Vandenesch, F. 1998, "Transmembrane topology and histidine protein kinase activity of AgrC, the *agr* signal receptor in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 28, no. 3, pp. 655-662.

Lina, G., Piemont, Y., Godail-Gamot, F., Bes, M., Peter, M. O., Gauduchon, V., Vandenesch, F., & Etienne, J. 1999, "Involvement of Panton-Valentine leukocidin-producing *Staphylococcus aureus* in primary skin infections and pneumonia", *Clinical Infectious Diseases*, vol. 29, no. 5, pp. 1128-1132.

Lina, G., Boutite, F., Tristan, A., Bes, M., Etienne, J., & Vandenesch, F. 2003, "Bacterial competition for human nasal cavity colonization: Role of staphylococcal *agr* alleles", *Applied and Environmental Microbiology*, vol. 69, no. 1, pp. 18-23.

Lindsay, J. A. & Foster, S. J. 1999, "Interactive regulatory pathways control virulence determinant production and stability in response to environmental conditions in *Staphylococcus aureus*", *Molecular and General Genetics*, vol. 262, no. 2, pp. 323-331.

Lindsay, J. A. & Holden, M. T. G. 2004, "*Staphylococcus aureus*: superbug, super genome?", *Trends in Microbiology*, vol. 12, no. 8, pp. 378-385.

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D., & Hinds, J. 2006, "Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes", *Journal of Bacteriology*, vol. 188, no. 2, pp. 669-676.

Lipshutz, R. J., Fodor, S. P. A., Gingeras, T. R., & Lockhart, D. J. 1999, "High density synthetic oligonucleotide arrays", *Nature Genetics*, vol. 21, pp. 20-24.

Lipsitch, M. 2001, "Microbiology - Bacterial population genetics and disease", *Science*, vol. 292, no. 5514, pp. 59-60.

Liu, Y. F., Manna, A., Li, R. G., Martin, W. E., Murphy, R. C., Cheung, A. L., & Zhang, G. Y. 2001, "Crystal structure of the SarR protein from *Staphylococcus aureus*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 12, pp. 6877-6882.

Lloyd, R. G. & Buckman, C. 1995, "Conjugational recombination in *Escherichia-Coli* - genetic-analysis of recombinant formation in Hfrxf(-) crosses", *Genetics*, vol. 139, no. 3, pp. 1123-1148.

Loughman, J. A., Fritz, S. A., Storch, G. A., & Hunstad, D. A. 2009, "Virulence gene expression in human community-acquired *Staphylococcus aureus* infection", *Journal of Infectious Diseases*, vol. 199, no. 3, pp. 294-301.

Low, D.E. 1997, Clinical microbiology: Issues in identification and susceptibility testing, pp233-252 *In* K. B. Crossley and G. L. Archer (ed.), The staphylococci in human disease. Churchill Livingstone, New York, NY.

Lowy, F. D. 1998, "Medical progress - *Staphylococcus aureus* infections", *New England Journal of Medicine*, vol. 339, no. 8, pp. 520-532.

Lowy, F. D. 2003, "Antimicrobial resistance: the example of *Staphylococcus aureus*", *Journal of Clinical Investigation*, vol. 111, no. 9, pp. 1265-1273.

Lucchini, S., Desiere, F., & Brussow, H. 1998, "The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts", *Virology*, vol. 246, no. 1, pp. 63-73.

Lucchini, S., Thompson, A., & Hinton, J. C. D. 2001, "Microarrays for microbiologists", *Microbiology*, vol. 147, pp. 1403-1414.

Luong, T. T., Ouyang, S., Bush, K., & Lee, C. Y. 2002, "Type 1 capsule genes of *Staphylococcus aureus* are carried in a staphylococcal cassette chromosome genetic element", *Journal of Bacteriology*, vol. 184, no. 13, pp. 3623-3629.

Luong, T. T. & Lee, C. Y. 2006, "The *arl* locus positively regulates *Staphylococcus aureus* type 5 capsule via an *mgrA*-dependent pathway", *Microbiology*, vol. 152, pp. 3123-3131.

Lyon, B. R. & Skurray, R. 1987, "Antimicrobial resistance of *Staphylococcus aureus* - genetic-basis", *Microbiological Reviews*, vol. 51, no. 1, pp. 88-134.

Ma, X. X., Ito, T., Tiensasitorn, C., Jamklang, M., Chongtrakool, P., Boyle-Vavra, S., Daum, R. S., & Hiramatsu, K. 2002, "Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains", *Antimicrobial Agents and Chemotherapy*, vol. 46, no. 4, pp. 1147-1152.

Ma, X. X., Ito, T., Chongtrakool, P., & Hiramatsu, K. 2006, "Predominance of clones carrying panton-valentine leukocidin genes among methicillin-resistant *Staphylococcus aureus* strains isolated in Japanese hospitals from 1979 to 1985", *Journal of Clinical Microbiology*, vol. 44, no. 12, pp. 4515-4527.

Ma, X. X., Ito, T., Kondo, Y., Cho, M., Yoshizawa, Y., Kaneko, J., Katai, A., Higashiide, M., Li, S., & Hiramatsu, K. 2008, "Two different Panton-Valentine leukocidin phage lineages predominate in Japan", *Journal of Clinical Microbiology*, vol. 46, no. 10, pp. 3246-3258.

Mack, D., Fischer, W., Krokotsch, A., Leopold, K., Hartmann, R., Egge, H., & Laufs, R. 1996, "The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear beta-1,6-linked glucosaminoglycan: Purification and structural analysis", *Journal of Bacteriology*, vol. 178, no. 1, pp. 175-183.

Maguire, G. P., Arthur, A. D., Boustead, P. J., Dwyer, B., & Currie, B. J. 1998, "Clinical experience and outcomes of community-acquired and nosocomial methicillin-

resistant *Staphylococcus aureus* in a northern Australian hospital", *Journal of Hospital Infection*, vol. 38, no. 4, pp. 273-281.

Mahmood, R. & Khan, S. A. 1990, "Role of upstream sequences in the expression of the staphylococcal enterotoxin-B gene", *Journal of Biological Chemistry*, vol. 265, no. 8, pp. 4652-4656.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J. J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M., & Spratt, B. G. 1998, "Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, pp. 3140-3145.

Maira-Litran, T., Kropec, A., Abeygunawardana, C., Joyce, J., Mark, G., Goldmann, D. A., & Pier, G. B. 2002, "Immunochemical properties of the staphylococcal poly-N-acetylglucosamine surface polysaccharide", *Infection and Immunity*, vol. 70, no. 8, pp. 4433-4440.

Mallonee, D. H., Glatz, B. A., & Pattee, P. A. 1982, "Chromosomal mapping of A gene affecting enterotoxin-A production in *Staphylococcus aureus*", *Applied and Environmental Microbiology*, vol. 43, no. 2, pp. 397-402.

Manna, A. & Cheung, A. L. 2001, "Characterization of *sarR*, a modulator of *sar* expression in *Staphylococcus aureus*", *Infection and Immunity*, vol. 69, no. 2, pp. 885-896.

Manna, A. C., Bayer, M. G., & Cheung, A. L. 1998, "Transcriptional analysis of different promoters in the *sar* locus in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 180, no. 15, pp. 3828-3836.

Manna, A. C. & Cheung, A. L. 2006a "Transcriptional regulation of the *agr* locus and the identification of DNA binding residues of the global regulatory protein SarR in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 60, no. 5, pp. 1289-1301.

Manna, A. C. & Cheung, A. L. 2006b "Expression of SarX, a negative regulator of *agr* and exoprotein synthesis, is activated by MgrA in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 188, no. 12, pp. 4288-4299.

Manna, A. C. & Ray, B. 2007, "Regulation and characterization of *rot* transcription in *Staphylococcus aureus*", *Microbiology*, vol. 153, pp. 1538-1545.

Maree, C. L., Daum, R. S., Boyle-Vavra, S., Matayoshi, K., & Miller, L. G. 2007, "Community-associated methicillin-resistant *Staphylococcus aureus* isolates causing healthcare-associated infections", *Emerging Infectious Diseases*, vol. 13, no. 2, pp. 236-242.

Marples, R. R., Richardson, J. F., & Desaxe, M. J. 1986, "Bacteriological characters of strains of *Staphylococcus aureus* submitted to a reference laboratory related to methicillin resistance", *Journal of Hygiene*, vol. 96, no. 2, pp. 217-223.

Marples, R. R. & Cooke, E. M. 1988, "Current problems with methicillin-resistant *Staphylococcus aureus*", *Journal of Hospital Infection*, vol. 11, no. 4, pp. 381-392.

Marples, R. R. & Reith, S. 1992, "Methicillin-resistant *Staphylococcus aureus* in England and Wales", *Commun Dis Rep CDR Rev*, vol. 2, no. 3, p. R25-R29.

Martinez, J. L. & Baquero, F. 2002, "Interactions among strategies associated with bacterial infection: Pathogenicity, epidemicity, and antibiotic resistance", *Clinical Microbiology Reviews* , vol. 15, no. 4, pp. 647-679.

Martins, A. & Cunha, M. D. R. S. 2007, "Methicillin resistance in *Staphylococcus aureus* and coagulase-negative staphylococci: Epidemiologilcal and molecular aspects", *Microbiology and Immunology*, vol. 51, no. 9, pp. 787-795.

Massey, R. C., Horsburgh, M. J., Lina, G., Hook, M., & Recker, M. 2006, "Opinion - The evolution and maintenance of virulence in *Staphylococcus aureus*: a role for host-to-host transmission?", *Nature Reviews Microbiology*, vol. 4, no. 12, pp. 953-958.

Matsumoto, Y., Kaito, C., Morishita, D., Kurokawa, K., & Sekimizu, K. 2007, "Regulation of exoprotein gene expression by the *Staphylococcus aureus* cvfB gene", *Infection and Immunity*, vol. 75, no. 4, pp. 1964-1972.

Mayer, L. W. 1988, "Use of plasmid profiles in epidemiologic surveillance of disease outbreaks and in tracing the transmission of antibiotic resistance", *Clinical Microbiology Reviews*, vol. 1, no. 2, pp. 228-243.

McDowell, P., Affas, Z., Reynolds, C., Holden, M. T. G., Wood, S. J., Saint, S., Cockayne, A., Hill, P. J., Dodd, C. E. R., Bycroft, B. W., Chan, W. C., & Williams, P. 2001, "Structure, activity and evolution of the group I thiolactone peptide quorum-

sensing system of *Staphylococcus aureus*", *Molecular Microbiology*, vol. 41, no. 2, pp. 503-512.

McGowan, J. E., Terry, P. M., Huang, T. S. R., Houk, C. L., & Davies, J. 1979, "Nosocomial infections with gentamicin-resistant *Staphylococcus aureus* - plasmid analysis as an epidemiologic tool", *Journal of Infectious Diseases*, vol. 140, no. 6, pp. 864-872.

McNamara, P. J., Milligan-Monroe, K. C., Khalili, S., & Proctor, R. A. 2000, "Identification, cloning, and initial characterization of *rot*, a locus encoding a regulator of virulence factor expression in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 182, no. 11, pp. 3197-3203.

Melles, D. C., van Leeuwen, W. B., Snijders, S. V., Horst-Kreft, D., Peeters, J. K., Verbrugh, H. A., & van Belkum, A. 2007, "Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*", *Journal of Microbiological Methods*, vol. 69, no. 2, pp. 371-375.

Melles, D. C., Schouls, L., Francois, P., Herzig, S., Verbrugh, H. A., van Belkum, A., & Schrenzel, J. 2009, "High-throughput typing of *Staphylococcus aureus* by amplified fragment length polymorphism (AFLP) or multi-locus variable number of tandem repeat analysis (MLVA) reveals consistent strain relatedness", *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 28, no. 1, pp. 39-45.

Menestrina, G., Dalla Serra, M., Comai, M., Coraiola, M., Viero, G., Werner, S., Colin, D. A., Monteil, H., & Prevost, G. 2003, "Ion channels and bacterial infection: the case of beta-barrel pore-forming protein toxins of *Staphylococcus aureus*", *Febs Letters*, vol. 552, no. 1, pp. 54-60.

Meyers, J. A., Sanchez, D., Elwell, L. P., & Falkow, S. 1976, "Simple agarose-gel electrophoretic method for identification and characterization of plasmid deoxyribonucleic-acid", *Journal of Bacteriology*, vol. 127, no. 3, pp. 1529-1537.

Michel, A., Agerer, F., Hauck, C. R., Herrmann, M., Ullrich, J., Hacker, J., & Ohlsen, K. 2006, "Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair", *Journal of Bacteriology*, vol. 188, no. 16, pp. 5783-5796.

Miles, A. A., Misra, S. S., & Irwin, J. O. 1938, "The estimation of the bactericidal power of the blood", *Journal of Hygiene*, vol. 38, no. 6, pp. 732-749.

Milheirico, C., Oliveira, D. C., & De Lencastre, H. 2007, "Multiplex PCR strategy for subtyping the staphylococcal cassette chromosome *mec* type IV in methicillin-resistant *Staphylococcus aureus*: 'SCC*mec* IV multiplex'", *Journal of Antimicrobial Chemotherapy*, vol. 60, no. 1, pp. 42-48.

Milkman, R., Raleigh, E. A., McKane, M., Cryderman, D., Bilodeau, P., & McWeeny, K. 1999, "Molecular evolution of the *Escherichia coli* chromosome. V. Recombination patterns among strains of diverse origin", *Genetics*, vol. 153, no. 2, pp. 539-554.

Milstone, A. M., Song, X., Beers, C., Berkowitz, I., Carroll, K. C., & Perl, T. M. 2008, "Unrecognized burden of methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococcus carriage in the pediatric intensive care unit", *Infection Control and Hospital Epidemiology*, vol. 29, no. 12, pp. 1174-1176.

Miragaia, M., Thomas, I. C., Couto, I., Enright, M. C., & De Lencastre, H. 2007, "Inferring a population structure for Staphylococcus epidermidis from multilocus sequence typing data", *Journal of Bacteriology*, vol. 189, no. 6, pp. 2540-2552.

Monecke, S., Slickers, P., Ellington, M. J., Kearns, A. M., & Ehricht, R. 2007, "High diversity of Panton-Valentine leukocidin-positive, methicillin-susceptible isolates of *Staphylococcus aureus* and implications for the evolution of community-associated methicillin-resistant S-aureus", *Clinical Microbiology and Infection*, vol. 13, pp. 1157-1164.

Mongkolrattanothai, K., Boyle, S., Murphy, T. V., & Daum, R. S. 2004, "Novel non-*mecA*-containing staphylococcal chromosomal cassette composite island containing *pbp4* and *tagF* genes in a commensal staphylococcal species: a possible reservoir for antibiotic resistance islands in *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 5, pp. 1823-1836.

Montgomery, C. P., Boyle-Vavra, S., & Adem, P. V. 2008, "Comparison of virulence in community-associated methicillin-resistant *Staphylococcus aureus* pulsotypes USA300 and USA400 in a rat model of pneumonia (vol 198, pg 561, 2008)", *Journal of Infectious Diseases*, vol. 198, no. 11, pp. 1725.

Moore, P. C. L. & Lindsay, J. A. 2002, "Molecular characterisation of the dominant UK methicillin-resistant *Staphylococcus aureus* strains, EMRSA-15 and EMRSA-16", *Journal of Medical Microbiology*, vol. 51, no. 6, pp. 516-521.

Moran, G. J., Krishnadasan, A., Gorwitz, R. J., Fosheim, G. E., McDougal, L. K., Carey, R. B., & Talan, D. A. 2006, "Methicillin-resistant S-aureus infections among patients in the emergency department", *New England Journal of Medicine*, vol. 355, no. 7, pp. 666-674.

Morfeldt, E., Taylor, D., Vongabain, A., & Arvidson, S. 1995, "Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII", *Embo Journal*, vol. 14, no. 18, pp. 4569-4577.

Morfeldt, E., Tegmark, K., & Arvidson, S. 1996, "Transcriptional control of the *agr*-dependent virulence gene regulator, RNAIII, in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 21, no. 6, pp. 1227-1237.

Murakami, K., Minamide, W., Wada, K., Nakamura, E., Teraoka, H., & Watanabe, S. 1991, "Identification of methicillin-resistant strains of staphylococci by polymerase chain-reaction", *Journal of Clinical Microbiology*, vol. 29, no. 10, pp. 2240-2244.

Murchan, S., Kaufmann, M. E., Deplano, A., de Ryck, R., Struelens, M., Zinn, C. E., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., El Solh, N., Cuny, C., Witte, W., Tassios, P. T., Legakis, N., van Leeuwen, W., van Belkum, A., Vindel, A., Laconcha, I., Garaizar, J., Haeggman, S., Olsson-Liljequist, B., Ransjo, U., Coombes, G., & Cookson, B. 2003, "Harmonization of pulsed-field gel electrophoresis protocols for

epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains", *Journal of Clinical Microbiology*, vol. 41, no. 4, pp. 1574-1585.

Murchan, S., Aucken, H. A., O'Neill, G. L., Ganner, M., & Cookson, B. D. 2004, "Emergence, spread, and characterization of phage variants of epidemic methicillin-resistant *Staphylococcus aureus* 16 in England and Wales", *Journal of Clinical Microbiology*, vol. 42, no. 11, pp. 5154-5160.

Mwangi, M. M., Wu, S. W., Zhou, Y. J., Sieradzki, K., De Lencastre, H., Richardson, P., Bruce, D., Rubin, E., Myers, E., Siggia, E. D., & Tomasz, A. 2007, "Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 22, pp. 9451-9456.

Naimi, T. S., Ledell, K. H., Boxrud, D. J., Groom, A. V., Steward, C. D., Johnson, S. K., Besser, J. M., O'Boyle, C., Danila, R. N., Cheek, J. E., Osterholm, M. T., Moore, K. A., & Smith, K. E. 2001, "Epidemiology and clonality of community-acquired methicillin-resistant *Staphylococcus aureus* in Minnesota, 1996-1998", *Clinical Infectious Diseases*, vol. 33, no. 7, pp. 990-996.

Naimi, T. S., Ledell, K. H., Como-Sabetti, K., Borchardt, S. M., Boxrud, D. J., Etienne, J., Johnson, S. K., Vandenesch, F., Fridkin, S., O'Boyle, C., Danila, R. N., & Lynfield, R. 2003, "Comparison of community- and health care-associated

methicillin-resistant *Staphylococcus aureus* infection", *Jama-Journal of the American Medical Association*, vol. 290, no. 22, pp. 2976-2984.

Narita, S., Kaneko, J., Chiba, J., Piemont, Y., Jarraud, S., Etienne, J., & Kamio, Y. 2001, "Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, phi SLT", *Gene*, vol. 268, no. 1-2, pp. 195-206.

Nelson, D. 2004, "Phage taxonomy: We agree to disagree", *Journal of Bacteriology*, vol. 186, no. 21, pp. 7029-7031.

NissenMeyer, J. & Nes, I. F. 1997, "Ribosomally synthesized antimicrobial peptides: Their function, structure, biogenesis, and mechanism of action", *Archives of Microbiology*, vol. 167, no. 2-3, pp. 67-77.

Noto, M. J., Kreiswirth, B. N., Monk, A. B., & Archer, G. L. 2008, "Gene acquisition at the insertion site for SCC*mec*, the genomic island conferring methicillin resistance in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 190, no. 4, pp. 1276-1283.

Novick, R. P., Ross, H. F., Projan, S. J., Kornblum, J., Kreiswirth, B., & Moghazeh, S. 1993, "Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule", *Embo Journal*, vol. 12, no. 10, pp. 3967-3975.

Novick, R. P., Projan, S. J., Kornblum, J., Ross, H. F., Ji, G., Kreiswirth, B., Vandenesch, F., & Moghazeh, S. 1995, "The *agr* P2 operon - an autocatalytic sensory

transduction system in *Staphylococcus aureus*", *Molecular & General Genetics*, vol. 248, no. 4, pp. 446-458.

Novick, R. P. 2000, Pathogenicity factors and their regulation, pp. 392-407. *In* V. A. Fischetti (ed.), Gram-positive pathogens. American Society for Microbiology, Washington, D.C.

Novick, R. P. & Jiang, D. R. 2003, "The staphylococcal *saeRS* system coordinates environmental signals with agr quorum sensing", *Microbiology*, vol. 149, pp. 2709-2717.

Nubel, U., Schmidt, P. M., Reiss, E., Bier, F., Beyer, W., & Naumann, D. 2004, "Oligonucleotide microarray for identification of *Bacillus anthracis* based on intergenic transcribed spacers in ribosomal DNA", *Fems Microbiology Letters*, vol. 240, no. 2, pp. 215-223.

O'Brien, F. G., Pearman, J. W., Gracey, M., Riley, T. V., & Grubb, W. B. 1999, "Community strain of methicillin-resistant *Staphylococcus aureus* involved in a hospital outbreak", *Journal of Clinical Microbiology*, vol. 37, no. 9, pp. 2858-2862.

O'Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G. F., & Ross, R. P. 2004, "Genome of Staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content", *Journal of Bacteriology*, vol. 186, no. 9, pp. 2862-2871.

O'Gara, J. P. & Humphreys, H. 2001, "*Staphylococcus epidermidis* biofilms: importance and implications", *Journal of Medical Microbiology*, vol. 50, no. 7, pp. 582-587.

O'Gara, J. P. 2007, "ica and beyond: biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*", *Fems Microbiology Letters*, vol. 270, no. 2, pp. 179-188.

O'Neill, A. J. & Chopra, I. 2006, "Molecular basis of fusB-mediated resistance to fusidic acid in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 59, no. 2, pp. 664-676.

O'Neill, G. L., Murchan, S., Gil-Setas, A., & Aucken, H. M. 2001, "Identification and characterization of phage variants of a strain of epidemic methicillin-resistant *Staphylococcus aureus* (EMRSA-15)", *Journal of Clinical Microbiology*, vol. 39, no. 4, pp. 1540-1548.

O'Toole, G., Kaplan, H. B., & Kolter, R. 2000, "Biofilm formation as microbial development", *Annual Review of Microbiology*, vol. 54, pp. 49-79.

Okuma, K., Iwakawa, K., Turnidge, J. D., Grubb, W. B., Bell, J. M., O'Brien, F. G., Coombs, G. W., Pearman, J. W., Tenover, F. C., Kapi, M., Tiensasitorn, C., Ito, T., & Hiramatsu, K. 2002, "Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community", *Journal of Clinical Microbiology*, vol. 40, no. 11, pp. 4289-4294.

Oliveira, D. C., Tomasz, A., & De Lencastre, H. 2002, "Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant Staphylocloccus aureus", *Lancet Infectious Diseases*, vol. 2, no. 3, pp. 180-189.

Oliveira, D. C., Milheirico, C., & De Lencastre, H. 2006, "Redefining a structural variant of staphylococcal cassette chromosome *mec*, SCC*mec* type VI", *Antimicrobial Agents and Chemotherapy*, vol. 50, no. 10, pp. 3457-3459.

Otter, J. A. & French, G. L. 2006, "Nosocomial transmission of community-associated meticillin-resistant *Staphylococcus aureus*: an emerging threat", *Lancet Infectious Diseases*, vol. 6, no. 12, pp. 753-755.

Otto, M., Sussmuth, R., Vuong, C., Jung, G., & Gotz, F. 1999, "Inhibition of virulence factor expression in *Staphylococcus aureus* by the *Staphylococcus epidermidis agr* pheromone and derivatives", *Febs Letters*, vol. 450, no. 3, pp. 257-262.

Pantucek, R., Doskar, J., Ruzickova, V., Kasparek, P., Oracova, E., Kvardova, V., & Rosypal, S. 2004, "Identification of bacteriophage types and their carriage in *Staphylococcus aureus*", *Archives of Virology*, vol. 149, no. 9, pp. 1689-1703.

Papakyriacou, H., Vaz, D., Simor, A., Louie, M., & McGavin, M. J. 2000, "Molecular analysis of the accessory gene regulator (*agr*) locus and balance of virulence factor expression in epidemic methicillin-resistant *Staphylococcus aureus*", *Journal of Infectious Diseases*, vol. 181, no. 3, pp. 990-1000.

Papia, G., Louie, M., Tralla, A., Johnson, C., Collins, V., & Simor, A. E. 1999, "Screening high-risk patients for methicillin-resistant *Staphylococcus aureus* on admission to the hospital: Is it cost effective?", *Infection Control and Hospital Epidemiology*, vol. 20, no. 7, pp. 473-477.

Pariza, M. W. & Iandolo, J. J. 1974, "Base ratio and deoxyribonucleic-acid homology studies of 6 *Staphylococcus aureus* typing bacteriophages", *Applied Microbiology*, vol. 27, no. 2, pp. 317-323.

Patel, A. H., Kornblum, J., Kreiswirth, B., Novick, R., & Foster, T. J. 1992, "Regulation of the protein A-encoding gene in *Staphylococcus aureus*", *Gene*, vol. 114, no. 1, pp. 25-34.

Peacock, S. J., Moore, C. E., Justice, A., Kantzanou, M., Story, L., Mackie, K., O'Neil, G., & Day, N. P. J. 2002, "Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*", *Infection and Immunity*, vol. 70, no. 9, pp. 4987-4996.

Peng, H. L., Novick, R. P., Kreiswirth, B., Kornblum, J., & Schlievert, P. 1988, "Cloning, characterization, and sequencing of an accessory gene regulator (*agr*) in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 170, no. 9, pp. 4365-4372.

Percival, S. L. & Bowler, P. G. 2004, "Biofilms and. their potential role in wound healing", *Wounds-A Compendium of Clinical Research and Practice*, vol. 16, no. 7, pp. 234-240.

Pereira, P. M., Filipe, S. R., Tomasz, A., & Pinho, M. G. 2007, "Fluorescence ratio imaging microscopy shows decreased access of vancomycin to cell wall synthetic sites in vancomycin-resistant *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 51, pp. 3627-3633.

Perichon, B. & Courvalin, P. 2009, "VanA-type vancomycin-resistant *Staphylococcus aureus*", *Antimicrob Agents Chemother*, vol. 53, no. 11, pp. 4580-4587.

Pi, B., Yu, M., Chen, Y., Yu, Y., & Li, L. 2009, "Distribution of the ACME-arcA gene among meticillin-resistant *Staphylococcus haemolyticus* and identification of a novel ccr allotype in ACME-arcA-positive isolates", *Journal of Medical Microbiology*, vol. 58, no. 6, pp. 731-736.

Piriz Duran, S., Kayser, F., & Berger-Baechi, B. 1996, "Impact of *sar* and *agr* on methicillin resistance in *Staphylococcus aureus*", *Fems Microbiology Letters*, vol. 141, no. 2-3, pp. 255-260.

Pitcher, D. G., Saunders, N. A., & Owen, R. J. 1989, "Rapid extraction of bacterial genomic DNA with guanidium thiocyanate", *Letters in Applied Microbiology*, vol. 8, no. 4, pp. 151-156.

Projan, S. J., and R. P. Novick. 1997, The molecular basis of pathogenicity, pp. 55-81. *In* K. B. Crossley and G. L. Archer (ed.), The staphylococci in human disease. Churchill Livingstone, New York, NY.

Prevost, G., Pottecher, B., Dahlet, M., Bientz, M., Mantz, J. M., & Piemont, Y. 1991, "Pulsed field gel-electrophoresis as a new epidemiologic tool for monitoring methicillin-resistant *Staphylococcus aureus* in an intensive-care unit", *Journal of Hospital Infection*, vol. 17, no. 4, pp. 255-269.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G. F., Desiere, F., & Brussow, H. 2002, "The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria", *Journal of Bacteriology*, vol. 184, no. 21, pp. 6026-6036.

Qi, W., Ender, M., O'Brien, F., Imhof, A., Ruef, C., McCallum, N., & Berger-Bachi, B. 2005, "Molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in Zurich, Switzerland (2003): Prevalence of type IVSCCmec and a new SCCmec element associated with isolates from intravenous drug users", *Journal of Clinical Microbiology*, vol. 43, no. 10, pp. 5164-5170.

Rechtin, T. M., Gillaspy, A. F., Schumacher, M. A., Brennan, R. G., Smeltzer, M. S., & Hurlburt, B. K. 1999, "Characterization of the SarA virulence gene regulator of *Staphylococcus aureus*", *Molecular Microbiology*, vol. 33, no. 2, pp. 307-316.

Recsei, P., Kreiswirth, B., Oreilly, M., Schlievert, P., Gruss, A., & Novick, R. P. 1986, "Regulation of exoprotein gene-expression in *Staphylococcus aureus* by *agr*", *Molecular & General Genetics*, vol. 202, no. 1, pp. 58-61.

Reed, S. B., Wesson, C. A., Liou, L. E., Trumble, W. R., Schlievert, P. M., Bohach, G. A., & Bayles, K. W. 2001, "Molecular characterization of a novel *Staphylococcus aureus* serine protease operon", *Infection and Immunity*, vol. 69, no. 3, pp. 1521-1527.

Richardson, J. F. & Reith, S. 1993, "Characterization of a strain of methicillin-resistant *Staphylococcus aureus* (EMRSA-15) by conventional and molecular methods", *Journal of Hospital Infection*, vol. 25, no. 1, pp. 45-52.

Roberts, C., Anderson, K. L., Murphy, E., Projan, S. J., Mounts, W., Hurlburt, B., Smeltzer, M., Overbeek, R., Disz, T., & Dunman, P. M. 2006, "Characterizing the effect of the *Staphylococcus aureus* virulence factor regulator, SarA, on log-phase mRNA half-lives", *Journal of Bacteriology*, vol. 188, no. 7, pp. 2593-2603.

Robinson, D. A. & Enright, M. C. 2003, "Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 12, pp. 3926-3934.

Robinson, D. A. & Enright, M. C. 2004, "Evolution of *Staphylococcus aureus* by large chromosomal replacements", *Journal of Bacteriology*, vol. 186, no. 4, pp. 1060-1064.

Robinson, D. A., Kearns, A. M., Holmes, A., Morrison, D., Grundmann, H., Edwards, G., O'Brien, F. G., Tenover, F. C., McDougal, L. K., Monk, A. B., & Enright, M. C. 2005a, "Re-emergence of early pandemic *Staphylococcus aureus* as a community-acquired meticillin-resistant clone", *Lancet*, vol. 365, no. 9466, pp. 1256-1258.

Robinson, D. A., Monk, A. B., Cooper, J. E., Feil, E. J., & Enright, M. C. 2005b, "Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 187, no. 24, pp. 8312-8321.

Rockett, J. C. 2003, "To confirm or not to confirm (microarray data) - that is the question", *Drug Discovery Today*, vol. 8, no. 8, pp. 343.

Rohde, H., Knobloch, J. K. M., Horstkotte, M. A., & Mack, D. 2001, "Correlation of *Staphylococcus aureus* icaADBC genotype and biofilm expression phenotype", *Journal of Clinical Microbiology*, vol. 39, no. 12, pp. 4595-4596.

Rohrer, S. & Berger-Bachi, B. 2003, "FemABX peptidyl transferases: a link between branched-chain cell wall peptide formation and beta-lactam resistance in gram-positive cocci", *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 3, pp. 837-846.

Rohwer, F. & Edwards, R. 2002, "The Phage Proteomic Tree: a genome-based taxonomy for phage", *Journal of Bacteriology*, vol. 184, no. 16, pp. 4529-4535.

Rolinson, G. N. 1998, "Forty years of beta-lactam research", *Journal of Antimicrobial Chemotherapy*, vol. 41, no. 6, pp. 589-603.

Rooijakkers, S. H. M., Ruyken, M., van Roon, J., van Kessel, K. P. M., van Strijp, J. A. G., & van Wamel, W. J. B. 2006, "Early expression of SCIN and CHIPS drives instant immune evasion by *Staphylococcus aureus*", *Cellular Microbiology*, vol. 8, no. 8, pp. 1282-1293.

Rossi, J., Bischoff, M., Wada, A., & Berger-Bachi, B. 2003, "*MsrR*, a putative cell envelope-associated element involved in *Staphylococcus aureus sarA* attenuation", *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 8, pp. 2558-2564.

Rouillard, J. M., Herbert, C. J., & Zuker, M. 2002, "OligoArray: genome-scale oligonucleotide design for microarrays", *Bioinformatics*, vol. 18, no. 3, pp. 486-487.

Roundtree, P. & Beard, M. 1958, "Further observations on infection with phase type 80 staphylococci in Australia", *Med Jour Australia Ii*, vol. 45, no. 24, pp. 789-795.

Rountree, P. M. & Freeman, B. M. 1955, "Infections caused by a particular phage type of *Staphylococcus aureus*", *Med Jour Australia Ii*, vol. 42, no. 5, pp. 157-161.

Sabersheikh S. S. 2003, "Studies on the expression of virulence-associated genes of *Staphylococcus aureus*. PhD thesis. Open University

Sabersheikh, S. S. & Saunders, N. A. 2004, "Quantification of virulence-associated gene transcripts in epidemic methicillin-resistant *Staphylococcus aureus* by real-time PCR", *Molecular and Cellular Probes*, vol. 18, no. 1, pp. 23-31.

Saha, B., Singh, A. K., Ghosh, A., & Bal, M. 2008, "Identification and characterization of a vancomycin-resistant *Staphylococcus aureus* isolated from Kolkata (South Asia)", *Journal of Medical Microbiology*, vol. 57, no. 1, pp. 72-79.

Said-Salim, B., Dunman, P. M., McAleese, F. M., Macapagal, D., Murphy, E., McNamara, P. J., Arvidson, S., Foster, T. J., Projan, S. J., & Kreiswirth, B. N. 2003,

"Global regulation of *Staphylococcus aureus* genes by rot", *Journal of Bacteriology*, vol. 185, no. 2, pp. 610-619.

Salgado, C. D. & Farr, B. M. 2003, "MRSA and VRE: Preventing patient-to-patient spread", *Infections in Medicine*, vol. 20, no. 4, pp. 194-200.

Sambanthamoorthy, K., Smeltzer, M. S., & Elasri, M. O. 2006, "Identification and characterization of *msa* (SA1233), a gene involved in expression of SarA and several virulence factors in *Staphylococcus aureus*", *Microbiology*, vol. 152, pp. 2559-2572.

Saravia-Otten, P., Muller, H. P., & Arvidson, S. 1997, "Transcription of *Staphylococcus aureus* fibronectin binding protein genes is negatively regulated by *agr* and an *agr*-independent mechanism", *Journal of Bacteriology*, vol. 179, no. 17, pp. 5259-5263.

Saris, P. E. J., Immonen, T., Reis, M., & Sahl, H. S. 1996, "Immunity to lantibiotics", *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, vol. 69, no. 2, pp. 151-159.

Saunders, N. A., Underwood, A., Kearns, A. M., & Hallas, G. 2004, "A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of *Staphylococcus aureus*", *Microbiology*, vol. 150, pp. 3763-3771.

Schaberg, D. R., Tompkins, L. S., & Falkow, S. 1981, "Use of agarose-gel electrophoresis of plasmid deoxyribonucleic-acid to fingerprint Gram-negative bacilli", *Journal of Clinical Microbiology*, vol. 13, no. 6, pp. 1105-1108.

Schaberg, D. R. & Zervos, M. 1986, "Plasmid analysis in the study of the epidemiology of nosocomial gram-positive cocci", *Reviews of Infectious Diseases*, vol. 8, no. 5, pp. 705-712.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. 1995, "Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray", *Science*, vol. 270, no. 5235, pp. 467-470.

Schleifer, K. H., Kilpperbalz, R., Fischer, U., Faller, A., & Endl, J. 1982, "Identification of *Micrococcus-candidus* atcc-14852 as a strain of *Staphylococcus epidermidis* and of *Micrococcus-caseolyticus* atcc-13548 and *Micrococcus-varians* atcc-29750 as members of a new species, *Staphylococcus-caseolyticus*", *International Journal of Systematic Bacteriology*, vol. 32, no. 1, pp. 15-20.

Schlichting, C., Branger, C., Fournier, J. M., Witte, W., Boutonnier, A., Wolz, C., Goullet, P., & Doring, G. 1993, "Typing of *Staphylococcus aureus* by pulsed-field gel-electrophoresis, zymotyping, capsular typing, and phage typing - resolution of clonal relationships", *Journal of Clinical Microbiology*, vol. 31, no. 2, pp. 227-232.

Schmidt, K. A., Manna, A. C., Gill, S., & Cheung, A. L. 2001, "SarT, a repressor of alpha-haemolysin in *Staphylococcus aureus*", *Infection and Immunity*, vol. 69, no. 8, pp. 4749-4758.

Schmidt, K. A., Donegan, N. P., Kwan, W. A., & Cheung, A. 2004, "Influences of sigma(B) and *agr* on expression of staphylococcal enterotoxin B (*seb*) in *Staphylococcus aureus*", *Canadian Journal of Microbiology*, vol. 50, no. 5, pp. 351-360.

Schumacher, M. A., Hurlburt, B. K., & Brennan, R. G. 2001, "Crystal structures of SarA, a pleiotropic regulator of virulence genes in S-aureus", *Nature*, vol. 409, no. 6817, pp. 215-219.

Scott, J. M. & Haldenwang, W. G. 1999, "Obg, an essential GTP binding protein of Bacillus subtilis, is necessary for stress activation of transcription factor sigma(B)", *Journal of Bacteriology*, vol. 181, no. 15, pp. 4653-4660.

Seidl, K., Stucki, M., Ruegg, M., Goerke, C., Wolz, C., Harris, L., Berger-Bachi, B., & Bischoff, M. 2006, "*Staphylococcus aureus* CcpA affects virulence determinant production and antibiotic resistance", *Antimicrobial Agents and Chemotherapy*, vol. 50, no. 4, pp. 1183-1194.

Severin, A., Tabei, K., Tenover, F., Chung, M., Clarke, N., & Tomasz, A. 2004, "High level oxacillin and vancomycin resistance and altered cell wall composition in *Staphylococcus aureus* carrying the staphylococcal *mecA* and the enterococcal *vanA* gene complex", *Journal of Biological Chemistry*, vol. 279, no. 5, pp. 3398-3407.

Severin, A., Wu, S. W., Tabei, K., & Tomasz, A. 2004, "Penicillin-binding protein 2 is essential for expression of high-level vancomycin resistance and cell wall synthesis in vancomycin-resistant *Staphylococcus aureus* carrying the enterococcal *vanA* gene complex", *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 12, pp. 4566-4573.

Seybold, U., Kourbatova, E. V., Johnson, J. G., Halvosa, S. J., Wang, Y. F., King, M. D., Ray, S. M., & Blumberg, H. M. 2006, "Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health

care-associated blood stream infections", *Clinical Infectious Diseases*, vol. 42, no. 5, pp. 647-656.

Sheehan, B. J., Foster, T. J., Dorman, C. J., Park, S., & Stewart, G. S. A. B. 1992, "Osmotic and growth-phase dependent regulation of the *eta*-gene of *Staphylococcus aureus* - a role for DNA supercoiling", *Molecular & General Genetics*, vol. 232, no. 1, pp. 49-57.

Shopsin, B., Mathema, B., Alcabes, P., Said-Salim, B., Lina, G., Matsuka, A., Martinez, J., & Kreiswirth, B. N. 2003, "Prevalence of *agr* specificity groups among *Staphylococcus aureus* strains colonizing children and their guardians", *Journal of Clinical Microbiology*, vol. 41, no. 1, pp. 456-459.

Shore, A., Rossney, A. S., Keane, C. T., Enright, M. C., & Coleman, D. C. 2005, "Seven novel variants of the staphylococcal chromosomal cassette mec in methicillin-resistant *Staphylococcus aureus* isolates from Ireland", *Antimicrobial Agents and Chemotherapy*, vol. 49, no. 5, pp. 2070-2083.

Sigurdardottir, B., Berg, J. V., Hu, J. F., Alamu, J., Mcnutt, L. A., Diekema, D. J., & Herwaldt, L. A. 2006, "Descriptive epidemiology and case-control study of patients colonized with vancomycin-resistant enterococcus and methicillin-resistant *Staphylococcus aureus*", *Infection Control and Hospital Epidemiology*, vol. 27, pp. 913-919.

Singh, A., Goering, R. V., Simjee, S., Foley, S. L., & Zervos, M. J. 2006, "Application of molecular techniques to the study of hospital infection", *Clinical Microbiology Reviews*, vol. 19, no. 3, pp. 512-530.

Skinner, D. & Keefer, C. S. 1941, "Significance of bacteremia caused by *Staphylococcus aureus* - A study of one hundred and twenty-two cases and a review of the literature concerned wiih experimental infection in animals", *Archives of Internal Medicine*, vol. 68, no. 5, pp. 851-875.

Smith, C. B., Noble, V., Bensch, R., Ahlin, P. A., Jacobson, J. A., & Latham, R. H. 1982, "Bacterial-flora of the vagina during the menstrual-cycle - findings in users of tampons, napkins, and sea sponges", *Annals of Internal Medicine*, vol. 96, no. 6, pp. 948-951.

Somerville, G. A., Chaussee, M. S., Morgan, C. I., Fitzgerald, J. R., Dorward, D. W., Reitzer, L. J., & Musser, J. M. 2002, "*Staphylococcus aureus* aconitase inactivation unexpectedly inhibits post-exponential-phase growth and enhances stationary-phase survival", *Infection and Immunity*, vol. 70, no. 11, pp. 6373-6382.

Southern, E. M. 1975, "Detection of specific sequences among DNA fragments separated by gel-electrophoresis", *Journal of Molecular Biology*, vol. 98, no. 3, pp. 503-517.

Stewart, P. R., Waldron, H. G., Lee, J. S., & Matthews, P. R. 1985, "Molecular relationships among serogroup-B bacteriophages of *Staphylococcus aureus*", *Journal of Virology*, vol. 55, no. 1, pp. 111-116.

Strommenger, B., Kettlitz, C., Weniger, T., Harmsen, D., Friedrich, A. W., & Witte, W. 2006, "Assignment of Staphylococcus isolates to groups by spa typing, SmaI macrorestriction analysis, and multilocus sequence typing", *Journal of Clinical Microbiology*, vol. 44, pp. 2533-2540.

Struelens, M. J. 1996, "Laboratory methods in the investigation of outbreaks of hospital-acquired infection", *Baillieres Clinical Infectious Diseases*, vol. 3, no. 2, pp. 267-288.

Sumby, P. & Waldor, M. K. 2003, "Transcription of the toxin genes present within the staphylococcal phage phi Sa3ms is intimately linked with the phage's life cycle", *Journal of Bacteriology*, vol. 185, no. 23, pp. 6841-6851.

Tacconelli, E., Venkataraman, L., De Girolami, P. C., & D'Agata, E. M. C. 2004, "Methicillin-resistant *Staphylococcus aureus* bacteraemia diagnosed at hospital admission: distinguishing between community-acquired versus healthcare-associated strains", *Journal of Antimicrobial Chemotherapy*, vol. 53, no. 3, pp. 474-479.

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L. Z., Takahashi, M., Ankai, A., Baba, S., Fukui, S., Lee, J. C., & Hiramatsu, K. 2005, "Whole-genome sequencing of Staphylococcus haemolyticus uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species", *Journal of Bacteriology*, vol. 187, no. 21, pp. 7292-7308.

Taneike, I., Otsuka, T., Dohmae, S., Saito, K., Ozaki, K., Takano, M., Higuchi, W., Takano, T., & Yamamoto, T. 2006, "Molecular nature of methicillin-resistant

Staphylococcus aureus derived from explosive nosocomial outbreaks of the 1980s in Japan", *Febs Letters*, vol. 580, no. 9, pp. 2323-2334.

Tegmark, K., Karlsson, A., & Arvidson, S. 2000, "Identification and characterization of SarH1, a new global regulator of virulence gene expression in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 37, no. 2, pp. 398-409.

Tenover, F. 1998, "Implications of VRSA", *Journal of Hospital Infection*, vol. 40, no. SUPPL. A, p. H1F.

Tenover, F. C. 1985, "Plasmid fingerprinting - a tool for bacterial strain identification and surveillance of nosocomial and community-acquired infections", *Clinics in Laboratory Medicine*, vol. 5, no. 3, pp. 413-436.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H., & Swaminathan, B. 1995, "Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel-electrophoresis - criteria for bacterial strain typing", *Journal of Clinical Microbiology*, vol. 33, no. 9, pp. 2233-2239.

Tenover, F. C. 2006a, "Mechanisms of antimicrobial resistance in bacteria", *American Journal of Medicine*, vol. 119, no. 6, p. S3-S10.

Tenover, F. C., McDougal, L. K., Goering, R. V., Killgore, G., Projan, S. J., Patel, J. B., & Dunman, P. M. 2006b, "Characterization of a strain of community-associated methicillin-resistant *Staphylococcus aureus* widely disseminated in the United States", *Journal of Clinical Microbiology*, vol. 44, no. 1, pp. 108-118.

Tenover, F. C. & Goering, R. V. 2009, "Methicillin-resistant *Staphylococcus aureus* strain USA300: origin and epidemiology", *Journal of Antimicrobial Chemotherapy*, vol. 63, no. 3, pp. 441-446.

Throup, J. P., Zappacosta, F., Lunsford, R. D., Annan, R. S., Carr, S. A., Lonsdale, J. T., Bryant, A. P., McDevitt, D., Rosenberg, M., & Burnham, M. K. R. 2001, "The srhSR gene pair from *Staphylococcus aureus*: Genomic and proteomic approaches to the identification and characterization of gene function", *Biochemistry*, vol. 40, no. 34, pp. 10392-10401.

Toledo-Arana, A., Merino, N., Vergara-Irigaray, M., Dearbouille, M., Penades, J. R., & Lasa, H. 2005, "*Staphylococcus aureus* develops an alternative, ica-independent biofilm in the absence of the *arlRS* two-component system", *Journal of Bacteriology*, vol. 187, no. 15, pp. 5318-5329.

Tremaine, M. T., Brockman, D. K., & Betley, M. J. 1993, "Staphylococcal enterotoxin-A gene (*sea*) expression is not affected by the accessory gene regulator (*agr*)", *Infection and Immunity*, vol. 61, no. 1, pp. 356-359.

Tristan, A., Ferry, T., Durand, G., Dauwalder, O., Bes, M., Lina, G., Vandenesch, F., & Etienne, J. 2007a, "Virulence determinants in community and hospital meticillin-resistant *Staphylococcus aureus*", *Journal of Hospital Infection*, vol. 65, pp. 105-109.

Tristan, A., Bes, M., Meugnier, H., Lina, G., Bozdogan, B., Courvalin, P., Reverdy, M. E., Enright, M. C., Vandenesch, F., & Etienne, J. 2007b, "Global distribution of

Panton-Valentine leukocidin-positive methicillin-resistant *Staphylococcus aureus*, 2006", *Emerging Infectious Diseases*, vol. 13, no. 4, pp. 594-600.

Troesch, A., Nguyen, H., Miyada, C. G., Desvarenne, S., Gingeras, T. R., Kaplan, P. M., Cros, P., & Mabilat, C. 1999, "Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays", *Journal of Clinical Microbiology*, vol. 37, no. 1, pp. 49-55.

Ubukata, K., Nonoguchi, R., Matsuhashi, M., & Konno, M. 1989, "Expression and inducibility in *Staphylococcus aureus* of the *mecA* gene, which encodes a methicillin-resistant *S. aureus* specific penicillin-binding protein", *Journal of Bacteriology*, vol. 171, no. 5, pp. 2882-2885.

Udo, E. E., Pearman, J. W., & Grubb, W. B. 1993, "Genetic-analysis of community isolates of methicillin-resistant *Staphylococcus aureus* in western-Australia", *Journal of Hospital Infection*, vol. 25, no. 2, pp. 97-108.

Valle, J., Toledo-Arana, A., Berasain, C., Ghigo, J. M., Amorena, B., Penades, J. R., & Lasa, I. 2003, "SarA and not sigma(B) is essential for biofilm development by *Staphylococcus aureus*", *Molecular Microbiology*, vol. 48, no. 4, pp. 1075-1087.

van Belkum, A., van Leeuwen, W., Kaufmann, M. E., Cookson, B., Forey, F., Etienne, J., Goering, R., Tenover, F., Steward, C., O'Brien, F., Grubb, W., Tassios, P., Legakis, N., Morvan, A., El Solh, N., de Ryck, R., Struelens, M., Salmenlinna, S., Vuopio-Varkila, J., Kooistra, M., Talens, A., Witte, W., & Verbrugh, H. 1998, "Assessment of resolution and intercenter reproducibility of results of genotyping

Staphylococcus aureus by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study", *Journal of Clinical Microbiology*, vol. 36, no. 6, pp. 1653-1659.

van Duijkeren, E., Ikawaty, R., Broekhuizen-Stins, M. J., Jansen, M. D., Spalburg, E. C., de Neeling, A. J., Allaart, J. G., van Nes, A., Wagenaar, J. A., & Fluit, A. C. 2008, "Transmission of methicillin-resistant *Staphylococcus aureus* strains between different kinds of pig farms", *Veterinary Microbiology*, vol. 126, no. 4, pp. 383-389.

van Ijperen, C. & Saunders, N. A. 2004, "Microarrays for bacterial typing: realistic hope or holy grail?", *Methods Mol Biol*, vol. 266, pp. 213-227.

van Leeuwen, W., van Nieuwenhuizen, W., Gijzen, C., Verbrugh, H., & van Belkum, A. 2000, "Population studies of methicillin-resistant and -sensitive *Staphylococcus aureus* strains reveal a lack of variability in the agrD gene, encoding a staphylococcal autoinducer peptide", *Journal of Bacteriology*, vol. 182, no. 20, pp. 5721-5729.

van Loo, I. H. M., Diederen, B. M. W., Savelkoul, P. H. M., Woudenberg, J. H. C., Roosendaal, R., van Belkum, A., Toom, N. L. D., Verhulst, C., van Keulen, P. H. J., & Kluytmans, J. A. J. W. 2007, "Methicillin-resistant *Staphylococcus aureus* in meat products, the Netherlands", *Emerging Infectious Diseases*, vol. 13, pp. 1753-1755.

Vandenesch, F., Kornblum, J., & Novick, R. P. 1991, "A temporal signal, independent of *agr*, is required for *hla* but not *spa* transcription in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 173, no. 20, pp. 6313-6320.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M. E., & Etienne, J. 2003, "Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: Worldwide emergence", *Emerging Infectious Diseases*, vol. 9, no. 8, pp. 978-984.

Vickers, A. A., Chopra, I., & O'Neill, A. J. 2007, "Intrinsic novobiocin resistance in Staphylococcus saprophyticus", *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 12, pp. 4484-4485.

Von Eiff, C., Peters, G., & Heilmann, C. 2002, "Pathogenesis of infections due to coagulase-negative staphylococci", *Lancet Infectious Diseases*, vol. 2, no. 11, pp. 677-685.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. 1995, "AFLP - a new technique for DNA-fingerprinting", *Nucleic Acids Research*, vol. 23, no. 21, pp. 4407-4414.

Voss, A., Loeffen, F., Bakker, J., Klaassen, C., & Wulf, M. 2005, "Methicillin-resistant *Staphylococcus aureus* in pig farming", *Emerging Infectious Diseases*, vol. 11, no. 12, pp. 1965-1966.

Voyich, J. A., Braughton, K. R., Sturdevant, D. E., Whitney, A. R., Said-Salim, B., Porcella, S. F., Long, R. D., Dorward, D. W., Gardner, D. J., Kreiswirth, B. N., Musser, J. M., & Deleo, F. R. 2005, "Insights into mechanisms used by *Staphylococcus*

aureus to avoid destruction by human neutrophils", *Journal of Immunology*, vol. 175, no. 6, pp. 3907-3919.

Voyich, J. M., Sturdevant, D. E., Braughton, K. R., Kobayashi, S. D., Lei, B. F., Virtaneva, K., Dorward, D. W., Musser, J. M., & Deleo, F. R. 2003, "Genome-wide protective response used by group A Streptococcus to evade destruction by human polymorphonuclear leukocytes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 4, pp. 1996-2001.

Vuong, C., Saenz, H. L., Gotz, F., & Otto, M. 2000, "Impact of the agr quorum-sensing system on adherence to polystyrene in *Staphylococcus aureus*", *Journal of Infectious Diseases*, vol. 182, no. 6, pp. 1688-1693.

Vuong, C., Kocianova, S., Voyich, J. M., Yao, Y. F., Fischer, E. R., Deleo, F. R., & Otto, M. 2004, "A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence", *Journal of Biological Chemistry*, vol. 279, no. 52, pp. 54881-54886.

Wagner, P. L. & Waldor, M. K. 2002, "Bacteriophage control of bacterial virulence", *Infection and Immunity*, vol. 70, no. 8, pp. 3985-3993.

Waldor, M. K. & Mekalanos, J. J. 1996, "Lysogenic conversion by a filamentous phage encoding cholera toxin", *Science*, vol. 272, no. 5270, pp. 1910-1914.

Wang, R., Braughton, K. R., Kretschmer, D., Bach, T. H. L., Queck, S. Y., Li, M., Kennedy, A. D., Dorward, D. W., Klebanoff, S. J., Peschel, A., Deleo, F. R., & Otto,

M. 2007, "Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA", *Nature Medicine*, vol. 13, no. 12, pp. 1510-1514.

Warner, J. B. & Lolkema, J. S. 2003, "CcpA-dependent carbon catabolite repression in bacteria", *Microbiology and Molecular Biology Reviews*, vol. 67, no. 4, pp. 475-490.

Wichelhaus, T. A., Schafer, V., Brade, V., & Boddinghaus, B. 2001, "Differential effect of rpoB mutations on antibacterial activities of rifampicin and KRM-1648 against *Staphylococcus aureus*", *Journal of Antimicrobial Chemotherapy*, vol. 47, no. 2, pp. 153-156.

Wielders, C. L. C., Vriens, M. R., Brisse, S., Graaf-Miltenburg, L. A. M., Troelstra, A., Fleer, A., Schmitz, F. J., Verhoef, J., & Fluit, A. C. 2001, "Evidence for *in-vivo* transfer of *mecA* DNA between strains of *Staphylococcus aureus*", *Lancet*, vol. 357, no. 9269, pp. 1674-1675.

Wildsmith, S. E. & Elcock, F. J. 2001, "Microarrays under the microscope", *Journal of Clinical Pathology-Molecular Pathology*, vol. 54, no. 1, pp. 8-16.

Wilson, M. 2001, "Bacterial biofilms and human disease", *Science Progress*, vol. 84, no. 3, pp. 235-254.

Wisplinghoff, H., Rosato, A. E., Enright, M. C., Noto, M., Craig, W., & Archer, G. L. 2003, "Related clones containing SCCmec type IV predominate among clinically significant Staphylococcus epidermidis isolates", *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 11, pp. 3574-3579.

Witney, A. A., Marsden, M., Holden, T. G., Stabler, R. A., Husain, S. E., Vass, J. K., Butcher, P. D., Hinds, J., & Lindsay, J. A. 2005, "Design, validation, and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics", *Applied and Environmental Microbiology*, vol. 71, no. 11, pp. 7504-7514.

Witte, W., Strommenger, B., Stanek, C., & Cuny, C. 2007, "Methicillin-resistant *Staphylococcus aureus* ST398 in humans and animals, Central Europe", *Emerging Infectious Diseases*, vol. 13, no. 2, pp. 255-258.

Wolz, C., Pohlmann-Dietze, P., Steinhuber, A., Chien, Y. T., Manna, A., van Wamel, W., & Cheung, A. 2000, "Agr-independent regulation of fibronectin-binding protein(s) by the regulatory locus sar in *Staphylococcus aureus*", *Molecular Microbiology*, vol. 36, no. 1, pp. 230-243.

Wu, S. W., deLencastre, H., & Tomasz, A. 1996a, "Sigma-B, a putative operon encoding alternate sigma factor of *Staphylococcus aureus* RNA polymerase: Molecular cloning and DNA sequencing", *Journal of Bacteriology*, vol. 178, no. 20, pp. 6036-6042.

Wu, S. W., Piscitelli, C., deLencastre, H., & Tomasz, A. 1996b, "Tracking the evolutionary origin of the methicillin resistance gene: Cloning and sequencing of a homologue of *mecA* from a methicillin-susceptible strain of *Staphylococcus sciuri*", *Microbial Drug Resistance-Mechanisms Epidemiology and Disease*, vol. 2, no. 4, pp. 435-441.

Wu, S. W., de Lancastre, H., & Tomasz, A. 2001, "Recruitment of the *mecA* gene homologue of *Staphylococcus sciuri* into a resistance determinant and expression of the resistant phenotype in *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 183, no. 8, pp. 2417-2424.

Wulf, M. & Voss, A. 2008, "MRSA in livestock animals - an epidemic waiting to happen?", *Clinical Microbiology and Infection*, vol. 14, no. 6, pp. 519-521.

Xiong, Y. Q., Bayer, A. S., Yeaman, M. R., van Wamel, W., Manna, A. C., & Cheung, A. L. 2004, "Impacts of SarA and *agr* in *Staphylococcus aureus* strain Newman on fibronectin-binding protein A gene expression and fibronectin adherence capacity in vitro and in experimental infective endocarditis", *Infection and Immunity*, vol. 72, no. 3, pp. 1832-1836.

Yarwood, J. A., Paquette, K. A., Tikh, I. B., Volper, E. A., & Greenberg, E. P. 2007, "Generation of virulence factor variants in *Staphylococcus aureus* biofilms", *Journal of Bacteriology*, vol. 189, pp. 7961-7967.

Yarwood, J. M., McCormick, J. K., & Schlievert, P. M. 2001, "Identification of a novel two-component regulatory system that acts in global regulation of virulence factors of *Staphylococcus aureus*", *Journal of Bacteriology*, vol. 183, no. 4, pp. 1113-1123.

Yarwood, J. M., Bartels, D. J., Volper, E. M., & Greenberg, E. P. 2004, "Quorum sensing in *Staphylococcus aureus* biofilms", *Journal of Bacteriology*, vol. 186, no. 6, pp. 1838-1850.

Yazdani, R., Oshaghi, M., Havayi, A., Pishva, E., Salehi, R., Sadeghizadeh, M., & Foroohesh, H. 2006, "Detection of icaAD gene and biofilm formation in *Staphylococcus aureus* isolates from wound infections", *Iranian Journal of Public Health*, vol. 35, no. 2, pp. 25-28.

You, Y. H., Fu, C. X., Zeng, X., Fang, D. S., Yan, X. M., Sun, B. C., Xiao, D., & Zhang, J. Z. 2008, "A novel DNA microarray for rapid diagnosis of enteropathogenic bacteria in stool specimens of patients with diarrhea", *Journal of Microbiological Methods*, vol. 75, no. 3, pp. 566-571.

Zetola, N., Francis, J. S., Nuermberger, E. L., & Bishai, W. R. 2005, "Community-acquired meticillin-resistant *Staphylococcus aureus*: an emerging threat", *Lancet Infectious Diseases*, vol. 5, no. 5, pp. 275-286.

Zhang, K. Y., McClure, J. A., Elsayed, S., Tan, J., & Conly, J. M. 2008, "Coexistence of Panton-Valentine leukocidin-positive and -negative community-associated methicillin-resistant *Staphylococcus aureus* USA400 sibling strains in a large Canadian health-care region", *Journal of Infectious Diseases*, vol. 197, no. 2, pp. 195-204.

Zhang, K. Y., McClure, J. A., Elsayed, S., & Conly, J. M. 2009, "Novel Staphylococcal Cassette Chromosome *mec* type, tentatively designated Type VIII, harboring class A *mec* and type 4 *ccr* gene complexes in a Canadian epidemic strain of methicillin-resistant *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, vol. 53, no. 2, pp. 531-540.

Zhu, W. M., Clark, N. C., McDougal, L. K., Hageman, J., McDonald, L. C., & Patel, J. B. 2008, "Vancomycin-resistant *Staphylococcus aureus* isolates associated with *Inc18-like vanA plasmids in Michigan*", *Antimicrobial Agents and Chemotherapy*, vol. 52, no. 2, pp. 452-457.

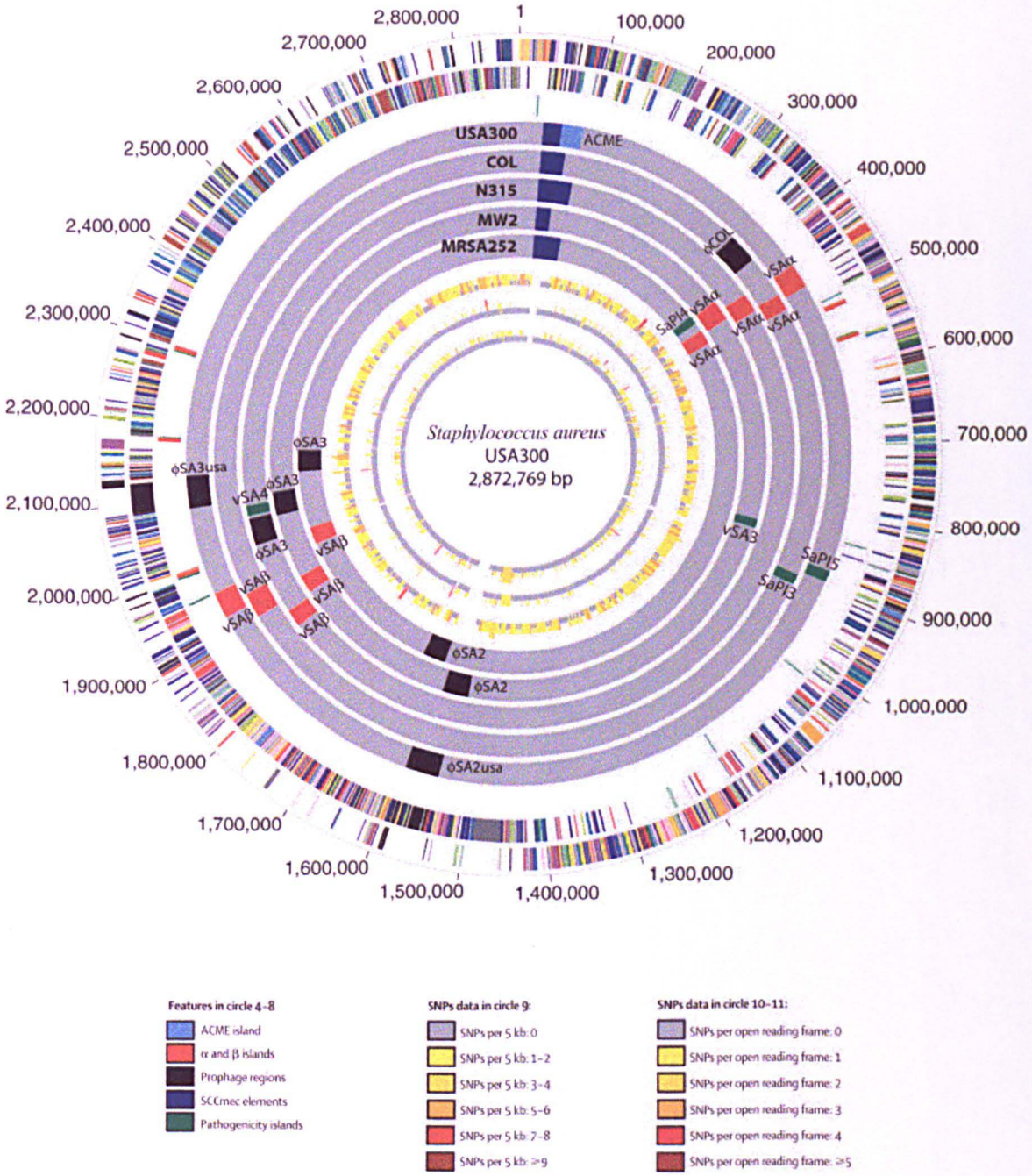
Ziebandt, A. K., Weber, H., Rudolph, J., Schmid, R., Hoper, D., Engelmann, S., & Hecker, M. 2001, "Extracellular proteins of *Staphylococcus aureus* and the role of SarA and sigma(B)", *Proteomics*, vol. 1, no. 4, pp. 480-493.

Ziebandt, A. K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M., & Engelmann, S. 2004, "The influence of *agr* and sigma(B) in growth phase dependent regulation of virulence factors in *Staphylococcus aureus*", *Proteomics*, vol. 4, no. 10, pp. 3034-3047.

Zou, D., Kaneko, J., Narita, S., & Kamio, Y. 2000, "Prophage, phi PV83-pro, carrying panton-valentine leukocidin genes, on the *Staphylococcus aureus* p83 chromosome: Comparative analysis of the genome structures of phi PV83-pro, phi PVL, phi 11, and other phages", *Bioscience Biotechnology and Biochemistry*, vol. 64, no. 12, pp. 2631-2643.

Appendix I: Chromosomal map of sequenced *S. aureus* strains

(Taken from Diep *et al.*, 2006)



Appendix II: Abbreviations

aa	amino-allyl
ACME	argenine catabolic mobile element
AFLP	amplified fragment length polymorphism
agr	accessory genome regulator
aha	amino hexylacrylamide
AIP	autoinducing peptide
Arl	autolysis related locus
CA-MRSA	community-associated methicillin-resistant *Staphylococcus aureus*
CA-MSSA	community-associated methicillin-sensitive *Staphylococcus aureus*
CC	clonal complex
CcpA	catabolite control protein a
CDFF	constant depth film fermentor
cDNA	complementary DNA
CFU	colony forming unit
CGH	comparative genome hybridisation
CHIPS	chemotaxis inhibitory protein of staphylococci
CMV	Cytomegalovirus
CNS	coagulase negative staphylococci
Cy	Cyanine
DLV	double locus variant
E. coli	*Escherichia coli*
EE	early exponential
egc	enterotoxin gene cluster
EMRSA	epidemic methicillin-resistant *Staphylococcus aureus*
GPhi	Genomiphi
HA-MRSA	healthcare-associated methicillin-resistant *Staphylococcus aureus*
HA-MSSA	healthcare-associated methicillin-sensitive *Staphylococcus aureus*
hla	haemolysin a/ alpha haemolysin
HPA	health protection agency
IVT	*In vitro* transcription
LE	late exponential
LHCAI	Laboratory of Healthcare Associated Infections
Luk	Leukocidin
MAD	median absolute deviation
MLEE	multilocus enzyme electrophoresis
MLST	multilocus sequence typing
MRCNS	methicillin-resistant coagulase negative Staphylococci
mRNA	messenger RNA
MRSA	methicillin-resistant *Staphylococcus aureus*
MRSE	methicillin-resistant *Staphylococcus aureus*
MSCRAMMS	microbial surface components recognising adhesive matrix molecules
MSSA	methicillin-sensitive *Staphylococcus aureus*
NCTC	National Collection of Type Cultures
NEB	New England Biolabs
OD	optical density
ON	overnight
PBP2a	penicillin binding protein 2a
PCR	Polymerase Chain Reaction
PE	planktonic exponential
PFGE	pulsed-field gel electrophoresis

PS	planktonic stationary
PSM	phenol soluble modulins
PTFE	polytetrafluoroethylene
PVL	Panton Valentine Leukocidin
RAP	RNAIII-activating protein
RIP	RNAIII inhibitory protein
rot	repressor of toxins
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase PCR
S. aureus	*Staphylococcus aureus*
Sae	Staphylococcal accessory element
SaPIn1	*Staphylococcus aureus* pathogenicity island 1
Sar	staphylococcal accessory regulator
SCC	staphylococcal chromosome cassette
SLST	single locus sequence typing
SLV	single locus variant
SNPs	single nucleotide polymorphisms
spaA	staphylococcal protein a
ST	sequence type
TRAP	target of RNAIII-activating protein
VISA	vancomycin intermediate-resistant *Staphylococcus aureus*
VNTR	variable number tandem repeat
VRSA	vancomycin resistant *Staphylococcus aureus*
WGA	whole genome amplification
ΦSa2	*Staphylococcus aureus* bacteriophage phi 2
ΦSa3	*Staphylococcus aureus* bacteriophage phi 3

Appendix IIII: Planned publications

Mohamed D, Saberesheikh S, Kearns A and Saunders N.

Comparison of healthcare-associated and community-associated *Staphylococcus aureus* using a truncated composite microarray.

Mohamed D, Saberesheikh S, Kearns A, Tediose T, Kane L and Saunders N.

Putative link between bacteriophage serotype and *Staphylococcus aureus* community-associated phenotype.

Mohamed D, Saberesheikh S, Kearns A and Saunders N.

Global regulation of virulence gene expression in a *Staphylococcus aureus agr* mutant.

Appendix IV: Supplementary data (see CD)